Zum Schätzproblem in Zeitreihen-Regressionen mit autokorrelierter Störvariable

Dissertation

zur Erlangung des Grades
eines Doktors der Wirtschaftswissenschaften
der Rechts- und Wirtschaftswissenschaftlichen Fakultät
der Universität des Saarlandes
(doctor rerum politicarum)

Erich-Otto Maschke

1975

Zum Schätzproblem in Zeitreihen-Regressionen mit autokorrelierter Störvariable

Dissertation

zur Erlangung des Grades
eines Doktors der Wirtschaftswissenschaften
der Rechts- und Wirtschaftswissenschaftlichen Fakultät
der Universität des Saarlandes
(doctor rerum politicarum)

Erich-Otto Maschke

Tag der Disputation : 24. Juni 1975
Dekan : Prof. Dr. Otto H. Poensgen
Erstberichterstatter : Prof. Dr. Hans Schneeweiß
Zweitberichterstatter: Prof. Dr. Volker Steinmetz

Inhalt

		Seite
0.	Fragestellung	1
1.	Schätzung in kleiner Stichprobe	4
1.1.	Schätzung im klassischen Modell	4
1.1.1.	Kleinstquadrat-Schätzung als beste lineare unverzerrte Schätzung	4
1.1.2.	Kollinearitätswirkung	5
1.2.	Schätzung bei autokorrelierter Störvariable	8
1.2.1.	Aitken-Schätzung	8
1.2.1.1.	Aitken-Schätzung als beste lineare unverzerrte Schätzung	8
1.2.1.2.	Darstellung der Varianz mittels Spektralverteilung	9
1.2.1.2.1.	Invertierbarer autoregressiver Störvariablen-Prozeß	9
1.2.1.2.2.	Stationärer Störvariablen-Prozeß mit positiver stetiger Spektraldichte	17
1.2.1.3.	Kollinearitätswirkung	22
1.2.1.4.	Beispiel	23
1.2.2.	Kleinstquadrat-Schätzung	24
1.2.2.1.	Varianz der Kleinstquadrat-Schätzung	24
1.2.2.2.	Varianz in Zeitreihenregressionen	25
1.2.2.3.	Varianz bei Regressoren mit linearem Trend	31
1.2.2.4.	Bedingungen für die Effizienz der Kleinstquadrat-Schätzung	37
1.2.2.5.	Eine Untergrenze der Varianz	43
1.2.3.	Schätzung nach prozeßkonformer Transformation des Modells	53
1.2.4.	Schätzung im Differenzen-Modell	58
1.2.5.	Approximative Aitken-Schätzung	63
1.2.6.	Effiziente Schätzung	65
2.	Schätzung in großer Stichprobe	69
2.1.	Voraussetzungen	69
2.2.	Asymptotische Varianz von Kleinstquadrat- und Aitken-Schätzung	72
2.3.	Bedingungen für die asymptotische Effizienz der Kleinstquadrat-Schätzung	77

		Seite
2.4.	Asymptotische Verteilung der Koeffizienten-schätzungen	79
2.5.	Konsistenz der Parameter-Schätzungen	87
2.6.	Trendbehaftete Regressoren	9o
2.6.1.	Polynomialer Trend	9o
2.6.2.	Exponentieller Trend	92
3.	Zusammenfassung	95
4.	Anhang	98
4.1.	Matrizen	98
4.2.	Summenformeln mit der Trendvariablen	1o1
4.3.	Varianz einer quadratischen Form	1o2
4.4.	Untergrenze des mittleren quadratischen Fehlers regulärer Schätzer	1o3
4.5.	Spektraldarstellung quasistationärer Prozesse	1o7

0. Fragestellung

Diese Arbeit beschäftigt sich mit dem Problem der Parameterschätzung in linearen Regressionen, deren Störvariable einem stationären Prozeß folgt. Im Vordergrund steht die Frage nach der Genauigkeit der Kleinstquadrat-Schätzung - verglichen mit der Genauigkeit alternativer Schätzverfahren bzw. verglichen mit ihrer Genauigkeit im autokorrelationsfreien Modell. Die Untersuchung ist im wesentlichen akzentuiert auf das Autokorrelationsproblem, wie es sich in der Zeitreihenökonometrie stellt.

In der Tat wird man dort häufig mit Autokorrelation der Störvariablen - genauer: positiver Autokorrelation - zu rechnen haben. Denn makroökonomische Aggregate sind in der Regel hoch auto- und reihenkorreliert, beispielsweise deshalb, weil sie einem gemeinsamen Trend unterliegen. Bringt man derartige Variablen in eine kausale Beziehung, dann wird man mit einer gewissen Berechtigung annehmen , daß die Störvariable dieser Beziehung - vorstellbar als (gewichtete) Summe von Variablen dieses Typs, wobei jeder Summand nur von vernachlässigbarem Einfluß auf die zu erklärende Variable ist - ebenfalls positive Autokorrelation aufweist. Des weiteren ist es nicht unplausibel und es erleichtert die Analyse ungemein, anzunehmen, daß bei laufendem Zeitindex - obwohl die in den Beziehungen verknüpften Variablen selbst gewöhnlich trendbehaftet sind - die Fehler in den Beziehungen sich (im Sinne einer konstanten Varianz) stabil verhalten und ihre Abhängigkeit voneinander (gemessen an der Kovarianz) mit wachsendem zeitlichen Abstand (und in Abhängigkeit nur von diesem) abnimmt.

Die Klasse der linearen Regressionen mit stationärem Störvariablen-Prozeß, wie sie hier als in gewisser Weise typisch für die zeitreihenökonometrische Praxis hingestellt wurden, wird für die folgende Untersuchung allerdings noch eingeschränkt auf nicht dynamische Modelle. Die Behandlung von Regressionen mit verzögerter endogener Variable erweist sich nämlich gerade, wenn man noch trendbehaftete echt exogene erklärende Variablen zulassen will, als außerordentlich schwierig.

Das behandelte Modell sieht also wie folgt aus:

(0.1)
$$
\begin{cases}
\text{(i)} \quad y = X\beta + u \qquad (\text{d.h. } y_t = x_t'\beta + u_t \quad (t = 1,\ldots,T)) \\
\qquad \text{mit von X unabhängigem } \beta, \text{ wobei X von vollem Spaltenrang ist und } \beta \text{ nicht eingeschränkt } (-\infty < \beta_i < \infty \quad (i = 1,\ldots,n)); \\
\text{(ii)} \quad E(u|X) = 0 \\
\text{(iii)} \quad E(uu'|X) = \Sigma = (\sigma(t-s))_{t,s=1,\ldots,T} \text{ und regulär sowie} \\
\qquad \text{unabhängig von X, wobei } 0 < \sigma^2 = \sigma(0) < \infty .
\end{cases}
$$

Dabei ist y der T-Vektor der erklärten Variablen, u der T-Vektor der Störvariablen, β der n-Vektor der Regressionskoeffizienten und X die (T x n)-Matrix mit den Zeilen $x_t' = (x_{1t},\ldots,x_{nt})$, die die Daten der erklärenden Variablen $x_1,\ldots,x_n$ enthält.

In Illustrationen wird später für gewöhnlich angenommen, daß die Autokovarianzen $\sigma(\tau)$ des u_t-Prozesses (zumindest für betragsmäßig kleines τ) positiv sind.
Die Schätzung in Zeitreihenregressionen vollzieht sich aber nicht allein vor dem Hintergrund einer möglicherweise positiv autokorrelierten Störvariable. Sie sieht sich vielmehr fast stets dem Problem hoher Kollinearität der Regressoren konfrontiert, die bereits im klassischen, autokorrelationsfreien Modell die Kleinstquadrat-Schätzung ungenau macht, bei positiv autokorrelierter Störvariable aber sich in Verbindung mit hoher Auto- und Reihenkorrelation der Regressoren noch weitaus fataler auswirkt (vgl. insbesondere 1.2.2.2., 1.2.2.3., aber auch 2.2. und 2.6.).

Nun hat für die kleine Stichprobe die geschilderte Aufgabenstellung einen hohen Komplexitätsgrad. Ihr wird deshalb im folgenden über längere Strecken durch Plausibilitätsbetrachtungen

begegnet. Für die große Stichprobe hingegen lassen sich unter recht allgemeinen Voraussetzungen präzise Argumente führen. Die Ergebnisse sind in Theoremen festgehalten. Soweit für die kleine Stichprobe Theoreme formuliert werden, handelt es sich dabei, von zwei Ausnahmen abgesehen, um Gegenstücke zu asymptotischen Resultaten über die Varianzen von Kleinstquadrat- und Aitken-Schätzung in Spektral-Repräsentation.

Der Anhang enthält zum Gebrauch im Text bestimmte mathematische Sätze. Diese sind größtenteils wohlbekannt und nur zur bequemeren Lektüre aufgeführt und hergeleitet.

1. Schätzung in kleiner Stichprobe

1.1. Schätzung im klassischen Modell

1.1.1. Kleinstquadrat-Schätzung als beste lineare unverzerrte Schätzung

Das klassische Modell ist dadurch gekennzeichnet, daß die Bedingung (O.1), (iii) $\Sigma = \sigma^2 I$ lautet, d.h. es ist $\sigma(t-s) = 0$ für $t \neq s$ $(t,s = 1,\ldots,T)$. Die Kleinstquadrat-Schätzung

$$\hat{\beta} = (X'X)^{-1}X'y = \beta + (X'X)^{-1}X'u \tag{1.1}$$

erfüllt unter dieser Bedingung

$$E(\hat{\beta}|X) = \beta \tag{1.2}$$

und hat die Varianz

$$V(\hat{\beta}|X) = \sigma^2(X'X)^{-1} \ . \tag{1.3}$$

Sie ist die bei gegebenem X eindeutig bestimmte beste lineare unverzerrte Schätzung. Denn jede lineare unverzerrte Schätzung $\check{\beta} = C'y$ (mit einer von y nicht abhängigen $(T \times n)$-Matrix C) muß $C'X = I$ erfüllen, also ist mit

$$P_X := I - X(X'X)^{-1}X' \tag{1.4}$$

$$V(\check{\beta}|X) - V(\hat{\beta}|X) = \sigma^2 C'C - \sigma^2(X'X)^{-1} = \sigma^2 C'P_X C = \sigma^2(P_X C)'P_X C \tag{1.5}$$

positiv-semidefinit, und das Gleichheitszeichen gilt genau dann, wenn $P_X C = 0$, d.h. $C = XH$ ist, und also wegen $C'X = I$ $C = X(X'X)^{-1}$.

Weiterhin ist mit

$$\hat{u} := y - X\hat{\beta} = P_X y = P_X u \tag{1.6}$$

(1.7) $\hat{\sigma}^2 := \hat{u}'\hat{u} / (T - n)$

eine (bei gegebenem X) erwartungstreue Schätzung von σ^2 und also

(1.8) $\hat{V}(\hat{\beta}|X) := \hat{\sigma}^2 (X'X)^{-1}$

eine (bei gegebenem X) erwartungstreue Schätzung von $V(\hat{\beta}|X)$.

1.1.2. Kollinearitätswirkung

Die Auswirkung von Kollinearität unter den Regressoren auf die Schätzgenauigkeit ersieht man für den Fall der inhomogenen Regression deutlich aus der bekannten Formel

$$V(\hat{\beta}_i|X) = \frac{\sigma^2}{Ts_i^2} \cdot \frac{1}{1-r^2_{i\cdot 2,\ldots,i-1,i+1,\ldots,n}} \qquad (i = 2,\ldots,n).^{1)}$$

Darin ist s_i^2 die (empirische) Varianz der Variablen x_i und $r^2_{i\cdot 2,\ldots,i-1,i+1,\ldots,n}$ das (empirische) multiple Bestimmheitsmaß zwischen x_i und den restlichen eigentlichen Regressoren. Gegenüber dem Fall völlig fehlender Kollinearität - i.e. unkorrelierter Regressoren - vervielfacht sich die Varianz der Kleinstquadrat-Schätzung $\hat{\beta}_i$ des eigentlichen Regressionskoeffizienten β_i (bei gleicher Varianz s_i) gerade um den Faktor $(1- r^2_{i\cdot 2,\ldots,i-1,i+1,\ldots,n})^{-1}$.

Da $r^2_{i\cdot 2,\ldots,i-1,i+1,\ldots,n} \geq r^2_{ik} \qquad (k \neq i,\ k = 2,\ldots,n)$ ist, gilt

(1.9) $$V(\hat{\beta}_i|X) \geq \sigma^2/Ts_i^2 \ (1 - \max_{\substack{k=2,\ldots,n \\ k\neq i}} r^2_{ik}) \geq \sigma^2/Ts_i^2 \ (1 - \min_{j,l=2,\ldots,n} r^2_{jl}),$$

wenn r_{jl} $(j,l = 2,\ldots,n)$ den einfachen Korrelationskoeffizienten zwischen x_j und x_l bezeichnet.

---

1) x_1 sei die Scheinvariable, $n \geq 3$

Nun sind für Zeitreihen makroökonomischer Aggregate die einfachen Korrelationskoeffizienten gewöhnlich (positiv und) recht hoch.[1] Entsprechend ungenau fallen nach (1.9) die Koeffizientenschätzungen in Zeitreihen-Regressionen aus[2] und zwar nicht etwa nur nach der Kleinstquadrat-Methode, sondern prinzipiell, da ja für jeden anderen linearen erwartungstreuen Schätzer $\check{\beta}$ $V(\check{\beta}_i|X) \geq V(\hat{\beta}_i|X)$ ist. Aus demselben Grund ist es (bei autokorrelationsfreier Störvariable) natürlich auch sinnlos, durch eine zeilenreguläre lineare Transformation der Datenreihen das Ausmaß der Kollinearität reduzieren zu wollen in der Hoffnung, daß im transformierten Modell eine bessere lineare Schätzung möglich ist.

Welche β_i (mit $c = e_i$, dem i-ten Einheitsvektor) und allgemeiner, welche Linearkombinationen $\alpha = c'\beta$ der Regressionskoeffizienten relativ (d.h. verglichen mit anderen Linearkombinationen) genau geschätzt werden können, zeigt die folgende Betrachtung (vgl. [Silvey]): Sind $\lambda_1,\ldots,\lambda_n$ die Eigenwerte von X'X, $v_1,\ldots,v_n$ die zugehörigen (orthonormalen) Eigenvektoren, so besitzt c die Darstellung $c = \sum_{i=1}^{n} q_i v_i$ mit $q_i = c'v_i$, $\sum_{i=1}^{n} q_i^2 = c'c$. Somit wird wegen $(X'X)^{-1}v_i = \lambda_i^{-1} v_i \quad (i = 1,\ldots,n)$

$$(1.10) \qquad \frac{1}{c'c} V(c'\hat{\beta}|X) = \sigma^2 \sum_{i=1}^{n} \frac{1}{\lambda_i} \frac{q_i^2}{\sum_{j=1}^{n} q_j^2} \quad .$$

Dies ist, abgesehen vom Faktor σ^2, ein gewogenes Mittel aus den Reziproken der Eigenwerte λ_i von X'X, jedes mit dem Gewicht des Anteils des Quadrats der senkrechten Projektion $q_i = c'v_i$ von c auf den zugehörigen Eigenvektor v_i am Quadrat der Länge des Vektors c versehen. Nach der Kleinstquadratmethode - und also

1) s. die Ergebnisse einer empirischen Studie von [Schulmann] zur Inter-, Auto-, Reihenkorrelation und Trendabhängigkeit makroökonomischer Niveau- und Differenzendaten (S.153 ff., 171ff., 185 ff.).

2) Auch das Absolutglied wird i.d.R. ungenau geschätzt, wie die Formel $V(\hat{\beta}_1|X) = \frac{1}{T}\sigma^2 + \bar{x}' V(\hat{\beta}_2,\ldots,_n|X)\bar{x}$ mit $\bar{x}' = (\bar{x}_2,\ldots,\bar{x}_n)$ $(\bar{x}_i := \sum_{t=1}^{T} x_{it}/T)$ und $\hat{\beta}_2,\ldots,_n = (\hat{\beta}_2,\ldots,\hat{\beta}_n)$ zeigt.

wiederum prinzipiell[1] - werden gewichtete Summen der Regressionskoeffizienten linear somit relativ am besten (schlechtesten) geschätzt, wenn der Vektor der Gewichte die Richtung des zum größten (kleinsten) Eigenwert von X'X gehörigen Eigenvektors hat.[2]

Schließlich sei noch erwähnt, daß hohe Kollinearität der Regressoren in der Regel auch eine starke stochastische Abhängigkeit der Schätzungen der verschiedenen Koeffizienten zur Folge hat; die Einflüsse der verschiedenen Regressoren lassen sich demnach nur schwer zuweisen.[3] [4]

---

1) Damit $\check{\alpha} = h'y$ (h unabhängig von y) erwartungstreue Schätzung von $\alpha = c'\beta$ ist, muß $X'h = c$ gelten, woraus $V(\check{\alpha}|X) - V(c'\hat{\beta}|X) = \sigma^2 h' P_X h$ (und weiterhin $h = X(X'X)^{-1}c$) folgt.

2) Hohe Kollinearität der Regressoren geht einher mit einem kleinen Wert von det X'X, d.h. mindestens ein Eigenwert von X'X ist klein; das ist ein heuristisches Argument dafür, daß sich in einem derartigen Fall bestimmte Linearkombinationen der Regressionskoeffizienten nur ungenau schätzen lassen.

3) Für den Korrelationskoeffizienten $\varrho(\hat{\beta}_i, \hat{\beta}_k)$ der Kleinstquadrat-Schätzungen der eigentlichen Regressionskoeffizienten β_i, β_k $(i \neq k)$ gilt nämlich (bei inhomogener Regression)

$$\varrho(\hat{\beta}_i, \hat{\beta}_k) = -\{r_{ik} + r_{i-(_{i,k}X)}\,{}_{(i,k}R)^{-1}\, r_{k-(_{i,k}X)}\} \Big/ \sqrt{1-r^2_{i\cdot(_{i,k}X)}}\,\sqrt{1-r^2_{k\cdot(_{i,k}X)}}.$$

Hierin sind die Größen $r^2_{\cdot}$ im Nenner multiple Bestimmtheitsmaße zwischen x_i bzw. x_k und jeweils der Gesamtheit der restlichen eigentlichen Regressoren x_j $(j \neq i,k)$; $r_{i-(_{i,k}X)}$ ist der Vektor der einfachen Korrelationskoeffizienten zwischen x_i und allen x_j mit $j \neq i,k$; entsprechend ist $r_{k-(_{i,k}X)}$ definiert; ${}_{i,k}R$ ist die Korrelationsmatrix der x_j $(j \neq i,k)$. Man gewinnt diese Formel durch die zweimalige Anwendung der Inversionsfromel (4.1.1) unter Benutzung von $V(\hat{\beta}|X) = T^{-1} S^{-1} R^{-1} S^{-1}$ (R : Korrelationsmatrix aller eigentlichen Regressoren, $S = \mathrm{diag}\,(s_2, \ldots, s_n)$), (Fortsetzung s. S.8)

4) Vgl. zum vorstehenden Abschnitt auch (1.66) und die dort anschließenden Erörterungen.

1.2. Schätzung bei autokorrelierter Störvariable

1.2.1. Aitken-Schätzung

1.2.1.1. Aitken-Schätzung als beste lineare unverzerrte Schätzung

Es gelte nunmehr das Modell (0.1) mit eventueller Ausnahme der Stationaritätsforderung $E(u_t\, u_s|X) = \sigma(t-s)$ $(t,s = 1,\ldots,T)$. Σ hat eine Zerlegung $\Sigma = \phi\phi'$, ϕ $(T \times T)$-regulär (beispielsweise mit $\phi = \Sigma^{1/2}$), und durch Transformation des Modells mit ϕ^{-1} entsteht das äquivalente klassische Modell

$$\phi^{-1}y = (\phi^{-1}X)\beta + (\phi^{-1}u) \tag{1.11}$$

mit $E(\phi^{-1}u|\phi^{-1}X) = E(\phi^{-1}u|X) = 0$ und

$E\left[(\phi^{-1}u)(\phi^{-1}u)'|\phi^{-1}X\right] = E\left[(\phi^{-1}u)(\phi^{-1}u)'|X\right] = I$. Jede lineare unverzerrte Schätzung $\check{\beta} = C'y$, $C'X = I$, im Modell (0.1) läßt sich auch als lineare unverzerrte Schätzung im Modell (1.11) auffassen, denn es ist $\check{\beta} = (\phi'C)'(\phi^{-1}y)$ und $(\phi'C)(\phi^{-1}X) = I$. Nach 1.1 ist daher jetzt die Aitken-Schätzung - d.i. die Kleinstquadrat-Schätzung im Modell (1.11) -

$$\hat{\beta} = (X'\Sigma^{-1}X)^{-1}\, X'\Sigma^{-1}y \tag{1.12}$$

die (bei gegebenem X) eindeutig bestimmte beste lineare unverzerrte Schätzung von β im Modell (0.1). Sie hat die Varianz

---

3) Fortsetzung von S.7

wobei man noch $r^2_{k\cdot(k}X) = r^2_{k\cdot(i,k}X) + (1-r^2_{k\cdot(i,k}X)\, r^2_{ki\cdot(i,k}X)$ (mit dem partiellen Korrelationskoeffizienten $r^2_{ki\cdot(i,k}X)$ zwischen x_k und x_i bei Bereinigung von den x_j, $j \neq i,k$ und multiplem Bestimmheitsmaß $r^2_{k\cdot(k}X)$) und die entsprechende Formel für $r^2_{i\cdot(i}X)$ und die Eigenschaft $r^2_{ki\cdot(i,k}X) = r^2_{ik\cdot(i,k}X)$ verwendet. Für $n = 3$ ist im übrigen gerade $\varrho(\hat{\beta}_2,\hat{\beta}_3) = -\, r_{23}$.

(1.13) $V(\hat{\beta}|X) = (X'\Sigma^{-1}X)^{-1}$.

Gleichfalls ist nun $\hat{\alpha} = c'\hat{\beta} = c'(X'\Sigma^{-1}X)^{-1} X'\Sigma^{-1}y$ die eindeutig bestimmte beste lineare Schätzung von $\alpha = c'\beta$.[1)]

Liegt des weiteren der homoskedastische Fall $E(u_t^2|X) = \sigma^2$ $(t = 1,\ldots,T)$ vor, dann ist mit

(1.14) $\Omega := {}^{1}/_{\sigma^2}\ \Sigma$

auch

(1.12') $\hat{\beta} = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1}y$.

Ferner ist dann

(1.15) $\hat{\sigma}^2 := y'(\Omega^{-1} - \Omega^{-1}X(X'\Omega^{-1}X)^{-1} X'\Omega^{-1})y / (T-n)$

eine (bei gegebenem X) erwartungstreue Schätzung von σ^2, sowie

(1.16) $\hat{V}(\hat{\beta}|X) := \hat{\sigma}^2(X'\Omega^{-1}X)^{-1}$

(bei gegebenem X) erwartungstreue Schätzung von $V(\hat{\beta}|X)$.

1.2.1.2. Darstellung der Varianz mittels Spektralverteilung

1.2.1.2.1. Invertierbarer autoregressiver Störvariablen-Prozeß

Für den Fall, daß die Störvariable u_t einem autoregressiven Prozeß folgt, gewinnt man eine interessante Darstellung von $V(\hat{\beta}|X)$ unter Benutzung der Spektralverteilungen des (theoretischen) u_t - wie des (empirischen) x_t-Prozesses.

---

1) S. Fußnote 1, S.7. Aus $X'h = c$ folgt nämlich $(\phi^{-1}X)'(\phi'h) = c$ und es ist $h'y = (\phi'h)'(\phi^{-1}y)$.

Zu diesem Zweck und zum späteren Gebrauch soll zunächst kurz die Spektralverteilungsfunktion einer positiv-semidefiniten Folge $\{M(\tau)\}_{\tau=\ldots,-1,0,+1,\ldots}$ reeller $(m \times m)$-Matrizen $M(\tau)$ charakterisiert werden. Eine solche Folge ist dadurch gekennzeichnet, daß für irgend endlich viele natürliche Zahlen $t_1 < \ldots < t_q$ die Matrix $\left[M(t_k - t_l)\right]_{k,l=1,\ldots,q}$ (d.i. die Matrix mit $M(t_k - t_l)$ im k-ten Zeilen- und l-ten Spaltenblock) positiv-semidefinit symmetrisch ist; gleichbedeutend damit ist Positiv-Semidefinitheit und Symmetrie der Matrix $\left[M(r-s)\right]_{r,s=1,\ldots,K}$ für jede natürliche Zahl K. Insbesondere muß also $M(-\tau) = (M(\tau))'$ sein und $M(0)$ bereits selbst positiv-semidefinit symmetrisch. Für jede solche Matrizenfolge gibt es nun eine eindeutig bestimmte i.a. komplexwertige $(m \times m)$-Matrixfunktion $\mathcal{F}(\lambda)$ $(-\infty < \lambda < \infty)$, die Spektralverteilungs-Funktion der Folge, mit:

(i) $\mathcal{F}(\lambda)$ ist überall (elementweise) rechtsstetig

(ii) $\mathcal{F}(\lambda)$ und die Zuwächse (d.h. die Matrizen $\mathcal{F}(\lambda_2) - \mathcal{F}(\lambda_1)$ für $\lambda_2 > \lambda_1$) sind positiv-semidefinit hermitesch

(iii) $\mathcal{F}(\lambda) = 0$ für $\lambda < -\pi$ [1]; $\mathcal{F}(\lambda) = M(0)$ für $\lambda \geq \pi$

(iv) $\int_{-\infty}^{\infty} e^{i\lambda\tau}\, d\mathcal{F}(\lambda) = \int_{-\pi}^{\pi} e^{i\lambda\tau}\, d\mathcal{F}(\lambda) = M(\tau) \qquad (\tau = 0, \pm 1, \ldots)$

Für jeden komplexen m-Vektor $c \neq 0$ ist $F_c(\lambda) := c^* \mathcal{F}(\lambda) c$ reell und (bis auf Normierung) eine Verteilungsfunktion; insbesondere trifft dies auf die Diagonalwerte $F_{jj}(\lambda)$ zu. Integration bzgl. $\mathcal{F}$ versteht sich elementweise; für die Diagonalelemente hat man ein Lebesgue-Stieltjes-Integral zu bilden, für die restlichen Elemente i.S. von (4.5.6) eine gewichtete Summe von Lebesgue-Stieltjes-Integralen. Die $\mathcal{F}$-integrierbaren Funktionen sind genau die bzgl. jedes F_c integrierbaren Funktionen.[2] - Das Integral in (iv) ist über das abgeschlossene Intervall $[-\pi, \pi]$ zu nehmen.

---

1) Statt dessen wird i.d.R. auch noch $\mathcal{F}(-\pi)=0$ verlangt; doch ist dann i.a. keine Rechtsstetigkeit mehr in $\lambda=-\pi$ gegeben.

2) Insbesondere ist jede beschränkte Baire-Funktion $\mathcal{F}$-integrierbar.

$\mathcal{F}(\lambda)$ ist genau dann stetig, wenn jedes $F_{jj}(\lambda)$ $(j = 1,\ldots,m)$ stetig ist; $\mathcal{F}(\lambda)$ hat also höchstens abzählbar viele Unstetigkeitsstellen, die außerdem symmetrisch zum Nullpunkt liegen. Überall existiert der linksseitige Limes $\mathcal{F}(\lambda-0)$. Die Unstetigkeitsstellen sind durchweg Sprünge, d.h. $\mathcal{F}(\lambda+0) - \mathcal{F}(\lambda-0)$ ist positiv-semidefinit. Insbesondere ist $\mathcal{F}(+0) - \mathcal{F}(-0) = \mathcal{R}\mathcal{F}(+0) - \mathcal{R}\mathcal{F}(-0)$ ($\mathcal{R}\mathcal{F}$ bezeichne den Realteil von $\mathcal{F}$).

Weiterhin gilt für $0 < \lambda_1 < \lambda_2$ (v) $\mathcal{F}(-\lambda_1) - \mathcal{F}(-\lambda_2) = \overline{\mathcal{F}(\lambda_2) - \mathcal{F}(\lambda_1)}$. [1]

Für ungerade Funktionen $g(\lambda)$ ist daher $\int_{-\pi}^{\pi} g(\lambda)\,d\mathcal{F} =$

$g(0)\{\mathcal{R}\mathcal{F}(+0) - \mathcal{R}\mathcal{F}(-0)\} + 2i \int_{0<\lambda\le\pi} g(\lambda)\,d\,\mathcal{I}m\mathcal{F}(\lambda)$ (wenn $\mathcal{I}m\mathcal{F}$ den Imaginärteil von $\mathcal{F}$ bezeichnet) und für gerade Funktionen $g(\lambda)$

$$\int_{-\pi}^{\pi} g(\lambda)\,d\mathcal{F}(\lambda) = g(0)\{\mathcal{R}\mathcal{F}(+0) - \mathcal{R}\mathcal{F}(-0)\} + 2\int_{0<\lambda\le\pi} g(\lambda)\,d\,\mathcal{R}\mathcal{F}(\lambda).$$

Beispielsweise ergibt sich also für das Integral (iv) :

$$\int_{-\pi}^{\pi} e^{i\lambda\tau}\,d\mathcal{F}(\lambda) = \mathcal{R}\mathcal{F}(+0) - \mathcal{R}\mathcal{F}(-0) + 2\int_{0<\lambda\le\pi} \cos\lambda\tau\,d\mathcal{R}\mathcal{F}(\lambda) - 2\int_{0<\lambda\le\pi} \sin\lambda\tau\,d\,\mathcal{I}m\mathcal{F}(\lambda).$$

Für $m = 1$ besteht $\mathcal{F}(\lambda)$ aus einem Element $F(\lambda)$; $F(\lambda)$ ist reell mit symmetrischen Zuwächsen (d.h. es ist $F(-\lambda_1) - F(-\lambda_2) = F(\lambda_2) - F(\lambda_1)$ für $0 < \lambda_1 < \lambda_2$ [1]) und (bis auf Normierung) eine Verteilungsfunktion. Hat $F(\lambda)$ eine Dichte $f(\lambda)$ bzgl. des Lebesgue-Maßes (d.h. ist $\int_A dF(\lambda) = \int_A f(\lambda)\,d\lambda$ für jede Lebesgue-meßbare Menge A), so gilt aufgrund der Symmetrie der Zuwächse von F $f(-\lambda) = f(\lambda)$ (Lebesgue - f.ü.). [2]

Ist nun der u_t-Prozeß (bei gegebenem X [3]) stationär, dann ist für jeden K-Vektor a $a'(\sigma(t-s))_{t,s=1,\ldots,K}a = V((u_1,\ldots,u_K)a|X) \ge 0$.

---

1) sofern λ_1, λ_2, $-\lambda_1$ und $-\lambda_2$ Stetigkeitsstellen sind

2) S. zu dem Vorstehenden etwa [Anderson], 7.3.

3) Man beachte, daß (0.1), (iii) nur über $E(u_t u_s|X)$ bei $t,s \le T$ aussagt, während der Prozeß aus der Gesamtheit der Variablen u_t, $t = 1,2,\ldots$ besteht. Die obige Forderung ist aber sicher-

(Fortsetzung s. S.12)

Der Prozeß besitzt also eine Spektralverteilungsfunktion $F(\lambda)$. Ist der Prozeß (bei gegebenem X) überdies autoregressiv, i.e.

$$(1.17)\qquad \sum_{r=0}^{p} \delta_r u_{t-r} = \varepsilon_t$$

$$\left[t = 1,2,\ldots;\quad \delta_o = 1,\quad \delta_p \neq 0;\quad E(\varepsilon_t|X) = 0 \quad (t = 1,2,\ldots);\right.$$

$$\left. E(\varepsilon_t \varepsilon_s|X) = \delta_{ts}\sigma_\varepsilon^2 \quad , \quad (t,s = 1,2,\ldots);\quad 0 < \sigma_\varepsilon^2 < \infty\right]$$

und invertierbar[1], dann hat er die überall positive und stetige Spektraldichte

$$(1.18)\qquad f(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi \left| \sum_{r=0}^{p} \delta_r e^{i\lambda r} \right|^2} \qquad (-\pi \leq \lambda \leq \pi)$$ [2].

Ferner hat auch die Folge $\{N^T(\tau)\}$ mit

$$(1.19)\qquad N^T(\tau) = \begin{cases} \sum_{t=1}^{T-\tau} x_{t+\tau}\, x_t' & , \quad 0 \leq \tau \leq T-1 \\ (N^T(-\tau))' & , \quad -(T-1) \leq \tau \leq -1 \\ 0 & , \quad |\tau| \geq T \end{cases}$$

---

3) Fortsetzung von S. 11

lich erfüllt, wenn für (fast) jede Realisation $(x_1, x_2, \ldots)$ des x_t-Prozesses $E(u_t u_s|x_1, x_2, \ldots) = \sigma(t-s)$ $(t,s = 1,2,\ldots)$ mit von der Realisation unabhängigem $\sigma(t-s)$ gilt; denn dann ist auch $E(u_t u_s|X) = E_{x_{T+1}, x_{T+2}, \ldots}\left[E(u_t u_s|x_1, x_2, \ldots)\right] = \sigma(t-s)$.

Es genügt jedenfalls nicht, nach dem Muster von (4.5.1) (Fs.s.S.13)

1) d.h. die Wurzeln der charakteristischen Gleichung $\sum_{r=0}^{p} \delta_r z^{p-r} = 0$ sind dem Betrage nach kleiner als 1

2) [Anderson], S.407

nach Lemma 4.5. eine (überall stetige) Spektralverteilungsfunktion $\mathcal{J}_T(\lambda)$. Weiterhin gilt mit der ((T-p) x T)-Matrix (für T > p)

$$(1.20)\qquad Q_\Delta := \begin{pmatrix} \delta_p & \cdots & \delta_0 & & 0 \\ & \delta_p & \cdots & \delta_0 & \\ & & \ddots & & \ddots \\ 0 & & & \delta_p & \cdots \delta_0 \end{pmatrix}$$

$$Q_\Delta u = \varepsilon_{(2)} := (\varepsilon_{p+1},\ldots,\varepsilon_T)'$$

und Zeile t von $Q_\Delta X$ lautet $x_{\Delta t} = \sum_{r=0}^{p} \delta_r x_{t+p-r}$ $\quad (t = 1,\ldots,T-p)$.

Definiert man noch mit

$$(1.21)\qquad D := \begin{pmatrix} \delta_0 & & 0 \\ \delta_1 & \delta_0 & \\ & \ddots & \\ \delta_{p-1} & \cdots & \delta_0 \end{pmatrix} \qquad 1)$$

$$(1.22)\qquad Q := \begin{bmatrix} D & & 0 \\ & Q_\Delta & \end{bmatrix},$$

so ist Q regulär und mit $u'_{(1)} := (u_1,\ldots,u_p)$

$$Q\,u = \begin{bmatrix} Du_{(1)} \\ \varepsilon_{(2)} \end{bmatrix} \quad \text{und daher mit } \Sigma_{11} := (E(u_t u_s|X))_{t,s=1,\ldots,p}$$

---

3) Fortsetzung von S.11 und 12

$\tilde{\sigma}(\tau) := \sigma(\tau)$ für $|\tau| \le T-1$, $\tilde{\sigma}(\tau) = 0$ für $|\tau| \ge T$ zu definieren, denn die so definierte Folge $\{\tilde{\sigma}(\tau)\}$ ist nicht notwendig positiv-semidefinit. - Vgl. aber zu diesem Punkt insbesondere noch S.51.

1) In der folgenden Betrachtung kann übrigens D auch einfach durch die Einheitsmatrix der Ordnung p ersetzt werden.

$$(1.23)\quad Q\,\Sigma\,Q' = V(Q\,u\,|\,X) = \begin{bmatrix} D\,\Sigma_{11}\,D' & 0 \\ 0 & \sigma_\varepsilon^2\,I \end{bmatrix},$$

denn aufgrund der Invertibilität des u_t-Prozesses ist $E(u_t \varepsilon_s | X) = 0$ für $t < s$. Daraus gewinnt man nunmehr die folgende nützliche Darstellung von Σ^{-1}:

$$(1.24)\quad \Sigma^{-1} = Q' \begin{bmatrix} D'^{-1}\,\Sigma_{11}^{-1}\,D^{-1} & 0 \\ 0 & 1/\sigma_\varepsilon^2 I \end{bmatrix} Q = \begin{bmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix} + 1/\sigma_\varepsilon^2\; Q_\Delta'\, Q_\Delta .$$

Mit $X_{(1)}' := [x_1, \ldots, x_p]$ und $X_\Delta := Q_\Delta X$ ist daher

$$(1.25)\quad X'\Sigma^{-1}X = X_{(1)}'\,\Sigma_{11}^{-1}\,X_{(1)} + 1/\sigma_\varepsilon^2\; X_\Delta'\, X_\Delta .$$

Weiterhin ist

$$X_\Delta'\, X_\Delta = \sum_{t=1}^{T-p} x_{\Delta t}\, x_{\Delta t}' = \sum_{t=1}^{T-p} \sum_{r,s=0}^{p} \delta_r \delta_s\, x_{t+p-r}\, x_{t+p-s}' =$$

$$= \sum_{r,s=0}^{p} \delta_r \delta_s \sum_{t=1}^{T-p} x_{t+p-r}\, x_{t+p-s}' =$$

$$= \sum_{r=0}^{p} \delta_r^2 \sum_{t=1}^{T} x_t x_t' + \sum_{\substack{r,s=0 \\ r<s}}^{p} \delta_r \delta_s \Big\{ \sum_{t=1}^{T-(s-r)} x_{t+s-r}\, x_t' + \sum_{t=1}^{T-(s-r)} x_t x_{t+s-r}' \Big\} + R_T ,\quad \text{also}$$

$$(1.26)\quad X_\Delta'\, X_\Delta = \sum_{r,s=0}^{p} \delta_r \delta_s\, N^T(s-r) + R_T$$

mit

$$(1.27)\quad R_T = -\sum_{r=0}^{p} \delta_r^2 \left\{ \sum_{t=1}^{p-r} x_t x_t' + \sum_{t=T-r+1}^{T} x_t x_t' \right\} -$$

$$- \sum_{\substack{r,s=0 \\ r<s}}^{p} \delta_r \delta_s \left[\sum_{t=1}^{p-s} \left\{ x_{t+s-r}\, x_t' + x_t x_{t+s-r}' \right\} + \right.$$

$$\left. + \sum_{t=T-s+1}^{T-s+r} \left\{ x_{t+s-r}\, x_t' + x_t x_{t+s-r}' \right\} \right] .$$

Nun gilt aber

$$\sum_{r,s=0}^{p} \delta_r \delta_s N^T(s-r) = \sum_{r,s=0}^{p} \delta_r \delta_s \int_{-\pi}^{\pi} e^{i\lambda(s-r)}\, d\mathcal{G}_T(\lambda),$$

$$= \int_{-\pi}^{\pi} \sum_{r,s=0}^{p} \delta_r \delta_s\, e^{i\lambda(s-r)}\, d\mathcal{G}_T(\lambda), \text{ also}$$

$$(1.28)\quad \frac{1}{\sigma_\varepsilon^2} \sum_{r,s=0}^{p} \delta_r \delta_s N^T(s-r) = \int_{-\pi}^{\pi} \frac{1}{\sigma_\varepsilon^2} \left| \sum_{r=0}^{p} \delta_r\, e^{i\lambda r} \right|^2 d\mathcal{G}_T(\lambda) =$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{f(\lambda)}\, d\mathcal{G}_T(\lambda).$$

In R_T treten nur die ersten sowie die letzten p x_t auf. Es schreibt sich in durchsichtigerer Weise als

$$(1.27')\quad R_T = -\sum_{r=0}^{p} \delta_r \left\{ \delta_r \sum_{t=1}^{p-r} x_t x_t' + \sum_{s=r+1}^{p} \delta_s \left(\sum_{t=1}^{p-r-(s-r)} x_{t+s-r} x_t' + \sum_{t=1}^{p-r-(s-r)} x_t x_{t+s-r}' \right) \right\}$$

$$- \sum_{s=0}^{p} \delta_s \left\{ \delta_s \sum_{t=T-s+1}^{T} x_t x_t' + \sum_{r=0}^{s-1} \delta_r \left(\sum_{t=T-s+1}^{T-(s-r)} x_{t+s-r}\, x_t' + \sum_{t=T-s+1}^{T-(s-r)} x_t' x_{t+s-r}' \right) \right\} .$$

Die hierin auftretenden Summen $\Sigma\, x_{t+s-r} x_t'$ und $\Sigma x_t\, x_{t+s-r}'$ stehen im Einklang mit der Definition (1.19), wenn man dieselbe auf die Systeme $\{x_1, \ldots, x_{p-r}\}$ rsp. $\{x_{T-s+1}, \ldots, x_T\}$ bezieht.

$X'_{(1)}\,\Sigma_{11}^{-1}\,X_{(1)} = \sum_{t,s=1}^{p} \sigma^{ts}\, x_t x_s'$ bereits zeigen, ist bei

$x_{it}^2 << \sum_{s=1}^{T} x_{is}^2 \qquad (t = 1,\ldots,T;\quad i = 1,\ldots,n)$ und $T >> p$

$\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{f(\lambda)}\, d\mathcal{G}_T(\lambda)$ eine recht gute Näherung von $X'\,\Sigma^{-1}\,X$ und also

(da die Inversion einer Matrix stetig in deren Elementen ist)

$2\pi \int_{-\pi}^{\pi} \left(\frac{1}{f(\lambda)}\, d\mathcal{G}_T(\lambda)\right)^{-1}$ eine recht gute Näherung von $(X'\Sigma^{-1}X)^{-1} = V(\hat{\beta}|X)$ [1)2)].

Die Darstellung (1.31) ist somit insbesondere unter asymptotischem Gesichtspunkt bedeutsam. Tatsächlich handelt es sich dabei auch um die Übertragung eines asymptotisch gültigen Ergebnisses von Grenander und Rosenblatt ([Grenander und Rosenblatt], S.240; s. 2.2.) auf die endliche Stichprobe. Interessanterweise gilt - wie Grenander und Rosenblatt zeigen - dieses Ergebnis nun nicht allein für invertierbare autoregressive Störvariablen-Prozesse, sondern grundsätzlich für stationäre Störvariablen-Prozesse mit stetiger positiver Spektraldichte $f(\lambda)$. [3)]

Zu jeder derartigen Spektraldichte $f(\lambda)$ gibt es nämlich für jedes $\eta > 0$ zwei Spektraldichten $f_1(\lambda)$, $f_2(\lambda)$ stationärer invertierbarer autoregressiver Prozesse mit

(1.32) $\quad 0 < f_1(\lambda) \le f(\lambda) \le f_2(\lambda);\quad \frac{1}{f_1(\lambda)} - \frac{1}{f_2(\lambda)} \le \eta \qquad (-\pi \le \lambda \le \pi).$

---

1) $\int_{-\pi}^{\pi} \frac{1}{f}\, d\,\mathcal{G}_T$ ist positiv-definit, da für jeden n-Vektor $c \neq 0$

$c' \int_{-\pi}^{\pi} \frac{1}{f}\, d\,\mathcal{G}_T\, c = \int_{-\pi}^{\pi} \frac{1}{f} d\,(c'\mathcal{G}_T\, c) \ge (\max_{-\pi\le\lambda\le\pi} f(\lambda))^{-1}\, c'\, N^T(0) c > 0$ ist.

2) Für eine etwas eingehendere Argumentation s. S.56 f.

3) Tatsächlich genügt sogar die stückweise Stetigkeit von $f(\lambda)$.

Bezeichnet man die zugehörigen Varianzen (bei festem T) mit Σ_1, Σ_2, so gilt wegen

$$c'\Sigma c = \sum_{t,s=1}^{T} c_t c_s \sigma(t-s) = \sum_{t,s=1} c_t c_s \int_{-\pi}^{\pi} e^{i\lambda(t-s)} f(\lambda) d\lambda = \int_{-\pi}^{\pi} \sum_{t,s=1} c_t c_s e^{i\lambda(t-s)} f(\lambda) d\lambda$$

$$= \int_{-\pi}^{\pi} \Big| \sum_{t=1}^{T} c_t e^{i\lambda t} \Big|^2 f(\lambda)\, d\lambda \qquad (c' = (c_1,\ldots,c_T))$$

(1.33) $\Sigma_1 \le \Sigma \le \Sigma_2$.

Σ_1, Σ_2 sind positiv-definit[1] und also $\Sigma_2^{-1} \le \Sigma^{-1} \le \Sigma_1^{-1}$, d.h.

(1.34) $X' \Sigma_2^{-1} X \le X' \Sigma^{-1} X \le X' \Sigma_1^{-1} X$.

Ferner ist wegen $c' \int_{-\pi}^{\pi} \frac{1}{f(\lambda)} d\mathfrak{g}_T(\lambda)\, c = \int_{-\pi}^{\pi} \frac{1}{f(\lambda)} d(c'\mathfrak{g}_T(\lambda)\, c)$

(1.35) $\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{f_2(\lambda)} d\mathfrak{g}_T(\lambda) \le \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{f(\lambda)} d\mathfrak{g}_T(\lambda) \le \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{f_1(\lambda)} d\mathfrak{g}_T(\lambda)$

wobei

(1.36) $0 \le \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{f_1(\lambda)} d\mathfrak{g}_T(\lambda) - \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{f_2(\lambda)} d\mathfrak{g}_T(\lambda) \le \frac{\eta}{2\pi} N^T(0)$

ist. Tatsächlich besagen die beiden letzten Ungleichungen, daß für die Elemente a_{ij} von $\frac{1}{2\pi} \int \frac{1}{f} d\mathfrak{g}_T - \frac{1}{2\pi} \int \frac{1}{f_2} d\mathfrak{g}_T$ bzw. b_{ij} von $\frac{1}{2\pi} \int \frac{1}{f_1} d\mathfrak{g}_T - \frac{1}{2\pi} \int \frac{1}{f} d\mathfrak{g}_T$

$0 \le a_{ij}, b_{ij} \le \frac{\eta}{2\pi} \sqrt{\sum_1^T x_{it}^2 \sum_1^T x_{jt}^2}$ gilt.

Für die asymptotische Betrachtung ist das bisher Gesagte (aufgrund von $X' \Sigma_1^{-1} X \doteq \frac{1}{2\pi} \int \frac{1}{f_1} d\mathfrak{g}_T$, $X' \Sigma_2^{-1} X \doteq \frac{1}{2\pi} \int \frac{1}{f_2} d\mathfrak{g}_T$ und (1.34) - (1.36) (vgl. 2.2.) ausreichend. Im vorliegenden

---

1) Vgl. Fn. 1, S.18

Die Matrizen

$$\sum_{t=1}^{p-r-(s-r)} (x_{t+s-r} \pm x_t)(x_{t+s-r} \pm x_t)' = \sum_{t=1}^{p-r-(s-r)} \{x_{t+s-r}x'_{t+s-r} + x_t x'_t \pm (x_{t+s-r}x'_t + x_t x'_{t+s-r})\}$$

sind positiv-semidefinit. Das hat zur Folge [1]

$$\sum_{t=1}^{p-r-(s-r)} x_{t+s-r}x'_t + \sum_{t=1}^{p-r-(s-r)} x_t x'_{t+s-r} \begin{cases} \geq -2\sum_{t=1}^{p-r} x_t x'_t \geq -2\sum_{t=1}^{p} x_t x'_t \\ \leq 2\sum_{t=1}^{p-r} x_t x'_t \leq 2\sum_{t=1}^{p} x_t x'_t \end{cases}$$

$$(0 \leq r < s \leq p).$$

Bei entsprechender Abschätzung im zweiten Summanden von R_T gewinnt man insgesamt [2]

$$(1.29) \quad -(\sum_{r=0}^{p} |\delta_r|)^2 \{\sum_{t=1}^{p} x_t x'_t + \sum_{t=T-p+1}^{T} x_t x'_t\} \leq R_T \leq (\sum_{r=0}^{p} |\delta_r|)^2 \{\sum_{t=1}^{p} x_t x'_t + \sum_{t=T-p+1}^{T} x_t x'_t\}.$$

Ebenfalls nur die ersten p x_t treten in dem Term $X'_{(1)} \Sigma_{11}^{-1} X_{(1)}$ auf. Bezeichnet man die Elemente von Σ_{11}^{-1} mit σ^{ts}, so gilt

$$\sigma^{ts} = \sigma^{st}, \quad |\sigma^{ts}| \leq \sqrt{\sigma^{tt}\,\sigma^{ss}} \leq \max_{t=1,\ldots,p} \sigma^{tt} \quad (t,s = 1,\ldots,p).$$

---

1) Hier und im folgenden soll die Relation $A \leq B$ zwischen reellen symmetrischen $(n \times n)$-Matrizen A,B bedeuten: $B - A$ ist positiv-semidefinit.- Die Relation ist transitiv. Aus $A \leq B$ folgt $A + C \leq B + C$ und $\lambda A \leq \lambda B$ $(\lambda > 0)$ bzw. $\lambda B \leq \lambda A$ $(\lambda < 0)$ für reelle Zahlen λ sowie $CAC' \leq CBC'$ für beliebige reelle $(m \times n)$-Matrizen C. Aus $0 < A \leq B$ (d.h.: A positiv-definit) folgt: B ist positiv-definit und $B^{-1} \leq A^{-1}$.

2) Unter Benutzung von $\sum_{r=0}^{p} |\delta_r|\{|\delta_r| + 2\sum_{s=r+1}^{p} |\delta_s|\} = \sum_{s=0}^{p} |\delta_s|\{|\delta_s| + 2\sum_{r=0}^{s-1} |\delta_r|\} = (\sum_{r=0}^{p} |\delta_r|)^2$

Spalte s von $X'_{(1)}\,\Sigma_{11}^{-1}$ lautet $\sum_{t=1}^{p} x_t\,\sigma^{ts}$. Daher ist

$$X'_{(1)}\,\Sigma_{11}^{-1}\,X_{(1)} = \sum_{s,t=1}^{p} \sigma^{ts}\,x_t x'_s = \sum_{t=1}^{p}\{\sigma^{tt}\,x_t x'_t + \sum_{s=t+1}^{p} \sigma^{ts}\,(x_t x'_s + x_s x'_t)\}$$

Nun ist aber $(x_t - x_s)(x_t - x_s)' \geq 0$, also $x_t x'_s + x_s x'_t \leq x_t x'_t + x_s x'_s$ $\times$

und ebenso $-(x_t x'_s + x_s x'_t) \leq x_t x'_t + x_s x'_s$; folglich ist

$$X'_{(1)}\Sigma_{11}^{-1}\,X_{(1)} \leq \sum_{t=1}^{p}\{\sigma^{tt}x_t x'_t + \sqrt{\sigma^{tt}}\sum_{s=t+1}^{p}\sqrt{\sigma^{ss}}\,x_t x'_t + \sqrt{\sigma^{tt}}\sum_{s=t+1}^{p}\sqrt{\sigma^{ss}}\,x_s x'_s\}\ ,\ \text{d.h.}$$

$$(1.30)\qquad 0 \leq X'_{(1)}\,\Sigma_{11}^{-1}\,X_{(1)} \leq 2p \max_{t=1,\ldots,p} \sigma^{tt} \sum_{t=1}^{p} x_t x'_t\ .\ ^{1)}$$

Insgesamt gilt also, wenn die Störvariable dem invertierbaren autoregressiven Prozeß $\sum_{r=0}^{p} \delta_r\,u_{t-r} = \varepsilon_t$ (wie in (1.17) angegeben) folgt:

Theorem 1.1. Es ist

$$(1.31)\qquad X'\,\Sigma^{-1}\,X = \tfrac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{f(\lambda)}\,d\mathfrak{g}_T(\lambda) + D_T$$

mit $D_T = X'_{(1)}\,\Sigma_{11}^{-1}\,X_{(1)} + \tfrac{1}{\sigma_\varepsilon^2}\,R_T$, wobei die Abschätzungen (1.30) und (1.29) gelten.

1.2.1.2.2. Stationärer Störvariablen-Prozeß mit positiver stetiger Spektraldichte

Wie die Abschätzungen lehren, wie aber auch (1.27) und die Darstellung

---

1) Tatsächlich besitzt die linke Seite eine Darstellung völlig analog zu der Darstellung (1.27') von R_T, wie man [Anderson], S.576, (85) und (1.24) entnehmen kann.

Fall hat man allerdings noch $D_T^{(i)} := X'\Sigma_i^{-1}X - \frac{1}{2\pi}\int \frac{1}{f_i}\, d\mathcal{J}_T$ $(i = 1,2)$ zu berücksichtigen. Erfüllen $f_1(\lambda)$, $f_2(\lambda)$ neben (1.32) noch

$$(1.37)\qquad f_2(\lambda) - f_1(\lambda) \le \eta \qquad (-\pi \le \lambda \le \pi)\ ,$$

so gilt für die zugehörigen Autokovarianzen

$$(1.38)\qquad |\sigma^{(2)}(\tau) - \sigma^{(1)}(\tau)| \le \int_{-\pi}^{\pi} |e^{i\lambda\tau}(f_2(\lambda) - f_1(\lambda))|\, d\lambda \le 2\pi\eta \qquad (\tau = 0, \pm 1, \dots).$$

Da die Matrixinversion eine stetige Operation ist, kann man also bei festem T durch entsprechende Wahl von $\eta > 0$ bei vorgegebenem $\varepsilon' > 0$ insgesamt erreichen, daß

$$\left| \frac{1}{2\pi}\int \frac{1}{f_1}\, d\mathcal{J}_T - \frac{1}{2\pi}\int \frac{1}{f_2}\, d\mathcal{J}_T \right| < \varepsilon', \quad |X'\Sigma_1^{-1}X - X'\Sigma_2^{-1}X| < \varepsilon',$$

$$|D_T^{(1)} - D_T^{(2)}| < 2\varepsilon' \quad \text{wird.}$$

Daß Funktionen $f_1(\lambda)$, $f_2(\lambda)$ mit (1.32) und (1.37) existieren, zeigt man wie folgt: Zu $0 < \eta' < (2\max f(\lambda))^{-1}$ gibt es für die reellen, positiven, in $-\pi \le \lambda \le \pi$ stetigen und geraden Funktionen $g_1(\lambda) = \frac{1}{f(\lambda)} + \frac{3}{4}\eta'$, $g_2(\lambda) = \frac{1}{f(\lambda)} - \frac{3}{4}\eta'$ nach dem Weierstraßschen Approximationssatz reellwertige trigonometrische Polynome $h_1(\lambda)$, $h_2(\lambda)$ der Form $h(\lambda) = \sum_{j=-p}^{p} c_j e^{i\lambda j}$, die ihrerseits gerade sind (d.h. es ist $c_j = c_{-j}$ reell, $j = 0,\dots,p$) und $|h_i(\lambda) - g_i(\lambda)| \le \eta'/4$ $(i = 1,2)$ $(-\pi \le \lambda \le \pi)$ erfüllen. Dann ist $\frac{1}{f(\lambda)} - \eta' < h_2(\lambda) < \frac{1}{f(\lambda)} - \eta'/2$ und $\frac{1}{f(\lambda)} + \eta'/2 < h_1(\lambda) < \frac{1}{f(\lambda)} + \eta'$ $(-\pi \le \lambda \le \pi)$. Setzt man $f_1(\lambda) = \frac{1}{h_1(\lambda)}$, $f_2(\lambda) = \frac{1}{h_2(\lambda)}$, so ist $0 < f_1(\lambda) < f(\lambda) < f_2(\lambda)$, $\frac{1}{f_1(\lambda)} - \frac{1}{f_2(\lambda)} < 2\eta'$ und bei geeigneter Wahl von η' neben $\frac{1}{f_1(\lambda)} - \frac{1}{f_2(\lambda)} < \eta$ auch $f_2(\lambda) - f_1(\lambda) < \eta$ $(-\pi \le \lambda \le \pi)$ (vgl. [Anderson], S.410 f.). Weiterhin ist

$$\int_{-\pi}^{\pi} e^{i\lambda\tau} h(\lambda)\, d\lambda = \int_{-\pi}^{\pi} e^{i\lambda\tau} \sum_{j=-p}^{p} c_j e^{-i\lambda j}\, d\lambda = \sum_{j=-p}^{p} c_j \int_{-\pi}^{\pi} e^{i\lambda(\tau - j)}\, d\lambda = \begin{cases} 2\pi c_\tau\ , & |\tau| \le p \\ 0\ , & |\tau| > p \end{cases}.$$

Definiert man die Folge $\tilde{\sigma}(\tau) := 2\pi c_\tau$ $(|\tau| \le p)$ und $\tilde{\sigma}(\tau) = 0$ $(|\tau| > p)$, so gilt also $\int_{-\pi}^{\pi} e^{i\lambda\tau} h(\lambda) d\lambda = \tilde{\sigma}(\tau)$, $\tilde{\sigma}(-\tau) = \tilde{\sigma}(\tau)$ und mit

$$d' = (d_1, \dots, d_K),\ K = 1,2,\dots,\ \sum_{t,s=1}^{K} d_t d_s \tilde{\sigma}(t-s) = \int_{-\pi}^{\pi} \Big| \sum_{t=1}^{K} d_t e^{i\lambda t} \Big|^2 h(\lambda) d\lambda \ge 0 .$$

Die Folge $\{\tilde{\sigma}(\tau)\}$ läßt sich somit als Autokovarianzfolge eines stationären Prozesses auffassen und zwar wegen $\tilde{\sigma}(\tau) = 0$ für $|\tau| > p$ als Autokovarianzfolge eines stationären Prozesses gleitender Durchschnitte der Ordnung p (s. [Anderson], S.224 f.). $h(\lambda)$ ist dessen Spektraldichte und also von der Form

$$h(\lambda) = \Big| \sum_{r=0}^{p} \delta_r e^{i\lambda r} \Big|^2 . \qquad 1/h(\lambda) = 1/\Big| \sum_{r=0}^{p} \delta_r e^{i\lambda r} \Big|^2 \text{ ist somit}$$

Spektraldichte eines stationären autoregressiven Prozesses

$$\sum_{r=0}^{p} \delta_r v_{t-r} = w_t \quad (E w_t = 0,\ E w_t w_s = \delta_{ts}\sigma_w^2,\ \sigma_w^2 = 1).^{1)}$$

Sind Wurzeln der charakteristischen Gleichung $\sum_{r=0}^{p} \delta_r z^{p-r} = 0$ dieses Prozesses dem Betrag nach größer als 1, so gibt es doch einen stationären autoregressiven Prozeß der Ordnung p mit derselben Spektraldichte, für den sämtliche Wurzeln der charakteristischen Gleichung betragsmäßig kleiner als 1 sind.[2)]

Insgesamt ist gezeigt

Theorem 1.2. Folgt die Störvariable einem stationären Prozeß mit positiver stetiger Spektraldichte $f(\lambda)$[3)], so gilt mit der gemäß (1.19) und Lemma 4.5. definierten zu $x_1, \dots, x_T$ gehörigen Spektralverteilungsfunktion $G_T(\lambda)$

---

1) Vgl. [Anderson, Sect. 5.2]. - Für $|z| = 1$, also $z = e^{i\lambda}$, ist $\Big| \sum_{r=0}^{p} \delta_r z^{p-r} \Big|^2 = h(\lambda) > 0$, d.h. die Wurzeln der charakteristischen Gleichung $\sum_{r=0}^{p} \delta_r z^{p-r} = 0$ sind betragsmäßig ungleich 1.

2) [Grenander und Rosenblatt] S.79 .

3) Über die Allgemeinheit dieser Voraussetzung s. 2.1.

(1.39) $X'\Sigma^{-1}X = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{f(\lambda)}\, d\mathcal{G}_T(\lambda) + D_T$.

Beide Summanden lassen sich mit beliebiger Genauigkeit durch die entsprechenden Terme der Zerlegung (1.31) von $X'\Sigma_a^{-1}X$ mit einer Varianz Σ_a eines invertierbaren autoregressiven Prozesses approximieren.

1.2.1.3. Kollinearitätswirkung

Es ist (vgl. Fn.1, S.18) bei stetigem positivem $f(\lambda)$

(1.40) $\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{f(\lambda)}\, d\mathcal{G}_T(\lambda) \leq (2\pi \min_{-\pi\leq\lambda\leq\pi} f(\lambda))^{-1}\, X'X$

und also

(1.41) $(\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{f(\lambda)}\, d\mathcal{G}_T(\lambda))^{-1} \geq 2\pi \min_{-\pi\leq\lambda\leq\pi} f(\lambda)\ (X'X)^{-1}$. [1)]

Liegt nun etwa der praxisrelevante Fall einer inhomogenen Regression mit linear oder polynomial trendbehafteten eigentlichen Regressoren und eines stationären autoregressiven Störvariablen-Prozesses niederer Ordnung (der ja in diesem Zusammenhang - s. S.21- in jedem Fall als invertierbar angesehen werden kann) vor - ein Fall, wie er bei nicht allzu kleinem Stichprobenumfang der auf S.18 geschilderten Situation entspricht - so wirkt sich, wie (1.41) zeigt, die unter Umständen i.d.R. recht hohe Kollinearität für die Aitken-Methode in Richtung einer ungenauen Schätzung aus.[2)]

---

1) Zur Illustration für den autoregressiven Prozeß:

$$2\pi f(\lambda) = \sigma_\varepsilon^2 / \left|\sum_{r=0}^{p} \delta_r e^{i\lambda r}\right|^2 \geq \sigma_\varepsilon^2 / \left(\sum_{r=0}^{p} |\delta_r|\right)^2 = \sum_{r=0}^{p} \delta_r \sigma(r) / \left(\sum_{r=0}^{p} |\delta_r|\right)^2 =$$

$$\sigma^2 \sum_{r=0}^{p} \delta_r \rho(r) / \left(\sum_{r=0}^{p} |\delta_r|\right)^2 \quad \text{mit Korrelationskoeffizienten } \rho(r).$$

2) Vgl. 1.1.2.

Diese Wirkung gilt damit natürlich auch prinzipiell für jedes lineare Schätzverfahren.

1.2.1.4. Beispiel

Zur Verdeutlichung insbesondere der Ergebnisse von 1.2.1.2.1. diene der autoregressive Prozeß erster Ordnung

$$(1.42)\quad u_t = \varrho\, u_{t-1} + \varepsilon_t \qquad (0 < \varrho < 1)\ .$$

Er hat die Spektraldichte

$$(1.43)\quad f(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi\,|1-\varrho e^{i\lambda}|^2} = \frac{\sigma^2(1-\varrho^2)}{2\pi\{(1-\varrho)^2+2\varrho(1-\cos\lambda)\}} \qquad (-\pi\le\lambda\le\pi).$$

Es ist $p = 1$ und $\Sigma_{11} = \sigma^2$. Ferner ist nach (1.27')

$$(1.44)\quad R_T = -\,x_1x_1' - \varrho^2 x_T x_T'\ .$$

Damit ist

$$(1.45)\quad D_T = X_{(1)}'\Sigma_{11}^{-1}X_{(1)} + \tfrac{1}{\sigma_\varepsilon^2}R_T = \tfrac{1}{\sigma^2}\,x_1x_1' + \tfrac{1}{\sigma_\varepsilon^2}R_T = -\frac{\varrho^2}{\sigma_\varepsilon^2}\,(x_1x_1' + x_Tx_T')\ .$$

Zur Kontrolle wird

$$\tfrac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{f(\lambda)}\,d\mathfrak{g}_T(\lambda) = \frac{1}{\sigma_\varepsilon^2}\int_{-\pi}^{\pi}(1+\varrho^2-2\varrho\cos\lambda)\,d\mathfrak{g}_T(\lambda)$$

gebildet. Nun ist

$$2\int_{-\pi}^{\pi}\cos\lambda\;d\mathfrak{g}_T(\lambda) = \int_{-\pi}^{\pi}(e^{i\lambda}+e^{-i\lambda})\,d\mathfrak{g}_T(\lambda) = N^T(1) + N^T(-1) \text{ und daher}$$

$$\tfrac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{f(\lambda)}\,d\mathfrak{g}_T(\lambda) = \tfrac{1}{\sigma_\varepsilon^2}\{(1+\varrho^2)\sum_{t=1}^{T}x_tx_t' - \varrho\sum_{t=1}^{T-1}x_{t+1}x_t' - \varrho\sum_{t=1}^{T-1}x_tx_{t+1}'\}\ .$$

Zusammen mit (1.45) ergibt dies tatsächlich das bekannte Ergebnis für $X'\Sigma^{-1}X$.

Im übrigen ist in diesem Fall ($\varrho > 0$) $\min f(\lambda) = \sigma^2(1-\varrho^2)/2\pi(1+\varrho^2)$, d.h. nach (1.41)

$$\left(\tfrac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{f(\lambda)}\,d\vartheta_T(\lambda)\right)^{-1} \geq \frac{1-\varrho^2}{1+\varrho^2}\,\sigma^2(X'X)^{-1}.$$

Außerdem ist D_T negativ-semidefinit und folglich auch

$$(X'\Sigma^{-1}X)^{-1} \geq \frac{1-\varrho^2}{1+\varrho^2}\,\sigma^2\,(X'X)^{-1} \quad . \tag{1.46}$$

1.2.2. Kleinstquadrat-Schätzung

1.2.2.1. Varianz der Kleinstquadrat-Schätzung

Die Varianz der Kleinstquadrat-Schätzung $\hat{\beta}$ ist

$$V(\hat{\beta}|X) = (X'X)^{-1}\,X'\,\Sigma\,X\,(X'X)^{-1}\ . \tag{1.47}$$

Mit den Elementen $\sigma_{ts} = \sigma(t-s)$ $\quad(t,s = 1,\ldots,T)$ von Σ lautet

Spalte s von $X'\Sigma$: $\sum_{t=1}^{T} x_t\,\sigma_{ts}$. Daher ist

$$X'\,\Sigma\,X = \sum_{t,s=1}^{T}\sigma_{ts}\,x_t x_s' = \sum_{t,s=1}^{T}\sigma(t-s)\,x_t x_s'$$

$$= \sigma(0)\sum_{t=1}^{T}x_t x_t' + \sum_{\tau=1}^{T-1}\sigma(\tau)\left(\sum_{t=1}^{T-\tau}x_{t+\tau}\,x_t' + \sum_{t=1}^{T-\tau}x_t\,x_{t+\tau}'\right)\ , \text{ also}$$

$$X'\,\Sigma\,X = \sum_{\tau=-(T-1)}^{T-1}\sigma(\tau)\,N^T(\tau) \tag{1.48}$$

mit $N^T(\tau)$ von (1.19).

Diese Darstellung eignet sich unmittelbar zur Verwendung der Spektralverteilungen, denn ist $F(\lambda)$ die Spektralverteilungsfunktion des u_t-Prozesses, so gilt

$$\sum_{\tau=-(T-1)}^{T-1} \sigma(\tau) N^T(\tau) = \sum_{\tau=-(T-1)}^{T-1} N^T(\tau) \int_{-\pi}^{\pi} e^{-i\lambda\tau} dF(\lambda) = \int_{-\pi}^{\pi} \sum_{\tau=-(T-1)}^{T-1} e^{-i\lambda\tau} N^T(\tau)\, dF(\lambda).$$ [1]

Mit der gemäß (4.5.2) definierten Spektraldichte $\mathcal{J}_T(\lambda)$ von $\mathcal{J}_T(\lambda)$ ist daher

(1.49) $$X' \Sigma X = 2\pi \int_{-\pi}^{\pi} \mathcal{J}_T(\lambda)\, dF(\lambda) .$$

Hat $F(\lambda)$ eine Dichte $f(\lambda)$, so gilt überdies

(1.49') $$X' \Sigma X = 2\pi \int_{-\pi}^{\pi} \mathcal{J}_T(\lambda)\, f(\lambda)\, d\lambda = 2\pi \int_{-\pi}^{\pi} f(\lambda)\, d\mathcal{J}_T(\lambda) .$$ [2]

1.2.2.2. Varianz in Zeitreihenregressionen

(1.49') lehrt, daß bei stetigem $f(\lambda)$ stets [3]

(1.50) $$2\pi \min_{-\pi\le\lambda\le\pi} f(\lambda)\, (X'X)^{-1} \le V(\hat{\beta}|X) \le 2\pi \max_{-\pi\le\lambda\le\pi} f(\lambda)\, (X'X)^{-1}$$

gilt. Eine andere Abschätzung ergibt sich aus (1.48) :

Es gilt nämlich 4)

(1.51) $$\pm\{N^T(\tau) + N^T(-\tau)\} = \pm\{N^T(\tau) + (N^T(\tau))'\} \le 2N^T(0) \qquad (\tau = 1,2,\ldots).$$

---

1) $\int_{-\pi}^{\pi} e^{i\lambda\tau} dF(\lambda) = \int_{-\pi}^{\pi} e^{-i\lambda\tau} dF(\lambda) = \int_{-\pi}^{\pi} \cos\lambda\tau\, dF(\lambda)$ (s.S.11).

2) Das zweite Gleichheitszeichen gilt nur, wenn $f(\lambda)$ tatsächlich bzgl. $\mathcal{J}_T(\lambda)$ integrierbar ist (s. dazu S.10); es gilt aber jedenfalls für (in $-\pi\le\lambda\le\pi$) stetiges $f(\lambda)$.

3) Zum Beweis s. Fn.1, S.18.

4) Das Beweisargument findet sich auf S.16.

Daraus folgt

$$(1.52)\quad X'\Sigma X = \sigma(0)N^T(0) + \sum_{\tau=1}^{T-1}\sigma(\tau)\{N^T(\tau) + N^T(-\tau)\} \leq \sum_{\tau=-(T-1)}^{T-1}|\sigma(\tau)|\, N^T(0)$$

und

$$(1.53)\quad V(\hat{\beta}|X) \leq \sum_{\tau=-(T-1)}^{T-1}|\sigma(\tau)|\quad (X'X)^{-1}.$$

Für den autoregressiven Störvariablen-Prozeß erster Ordnung (1.42) lautet, da in diesem Fall $\sigma(\tau) = \sigma^2\,\rho^{|\tau|}$ ist, diese Ungleichung beispielsweise

$$(1.54)\quad V(\hat{\beta}|X) \leq \sigma^2(1 + 2\sum_{\tau=1}^{T-1}\rho^\tau)(X'X)^{-1} = \frac{\sigma^2}{1-\rho}\,(1+\rho - 2\rho^T)(X'X)^{-1}.$$

Andererseits ist nach (1.43) $2\pi \max_{-\pi\leq\lambda<\pi} f(\lambda) = \sigma^2\,\frac{1+\rho}{1-\rho}$.

Für großes T besagen in diesem Fall (1.53) und die rechte Ungleichung von (1.50) dasselbe.

Dies ist bei positiver Autokorrelation der u_t (i.e. $\sigma(\tau) \geq 0$, $\tau=1,2,\ldots$) allgemein so, wenn deren Stärke für $\tau\to\infty$ so stark abklingt, daß nicht nur $\sigma(\tau) \to 0$ strebt, sondern auch $\sum_{\tau=-\infty}^{\infty}\sigma(\tau)$ (oder, was dasselbe ist, $\sum_{\tau=1}^{\infty}\sigma(\tau)$) konvergiert. Dann hat der u_t-Prozeß nämlich (s. [Anderson], S.380) die Spektraldichte $^1/_{2\pi}\sum_{\tau=-\infty}^{\infty}\sigma(\tau)\cos\lambda\tau$, die ihr Maximum in $\lambda=0$ mit dem Wert $^1/_{2\pi}\sum_{\tau=-\infty}^{\infty}\sigma(\tau)$ annimmt.

In dem für die Zeitreihen-Ökonometrie wichtigen Fall (polynom-) trendbehafteter (eigentlicher) Regressoren wird für $T\to\infty$ für einen derartigen u_t-Prozeß die angegebene Obergrenze von $V(\hat{\beta})$ - und übrigens auch von der Varianz des Aitken-Schätzers - tatsächlich angenommen. Der Grund ist der, daß in diesem Fall die für jedes $\tau = 0,1,2,\ldots$ gebildete(mit demselben Normalisie-

rungsfaktor versehene) Folge $\{N^T(\tau)\}_{T=\tau+1,\tau+2,\ldots}$ denselben Limes N(O) hat, der aufgrund der (absoluten) Konvergenz von $\sum \sigma(\tau)$ und der in T und τ gleichmäßigen Beschränkung der $N^T(\tau)$ (s. (1.51)) beim Übergang $T\to\infty$ in (1.48) eingetragen werden kann [1)]

Bei kleinem Stichprobenumfang T sind die

$N^T(\tau) = \sum_{t=1}^{T-\tau} x_{t+\tau} x_t'$ (für $\tau \geq 0$), mit denen die $\sigma(\tau)$ ja in (1.48) gewogen werden, hingegen durchaus verschieden und zwar auch, soweit sie einen bedeutenden Beitrag zur Summe leisten (was etwa - bei $\sigma(0) > \sigma(1) > \sigma(2) > \ldots$ und $x_{i\,t}^2 << \sum_{s=1}^{T} x_{i\,s}^2$, $i = 1,\ldots,n$ - für $|\tau| \leq \tau_0(T)$ der Fall sein möge). $V(\hat{\beta})$ nimmt dann die angegebene Obergrenze - die ja i.d.R. (vgl. (1.54)) eine beträchtliche Vergrößerung der Schätzungenauigkeit gegenüber dem klassischen Modell darstellt - nicht an.

Dennoch wird man im Zeitreihenmodell im allgemeinen mit einer generellen Verschlechterung der Schätzgenauigkeit gegenüber dem autokorrelationsfreien Fall zu rechnen haben. Diese Behauptung ist gleichwertig mit der Aussage

(1.55) $X' \Sigma X \geq \sigma^2 X'X$

oder, was dasselbe ist:

(1.55') $\sum_{\tau=1}^{T-1} \sigma(\tau)\ (N^T(\tau) + N^T(-\tau))$ ist positiv-semidefinit.

Zwar kommt, wie (1.50) zeigt, Kollinearität der Regressoren in der Varianz der Kleinstquadrat-Schätzung (bei positiv stetiger Spektraldichte des u_t-Prozesses) auf jeden Fall (mindestens implizit) zur Geltung. Die angeführte Behauptung besagt nun aber, daß erstens Kollinearität der Regressoren nicht nur

---

1) s. 2.6.1. - Natürlich gilt dies nur für eine normalisierte Varianz, da $V(\hat{\beta})$ selbst gegen null strebt.

in vielleicht (durch einen Faktor $\leq \sigma^2$) abgeschwächter Form, sondern voll zur Geltung kommt,[1] und daß zweitens darüberhinaus in Verbindung mit positiver Autokorrelation der Störvariablen eine in den $N^T(\tau)$ sich manifestierende positive Autokorrelation der Beobachtungsreihen x_t [2] die Schätzungenauigkeit noch vergrößert.

Zur Begründung von (1.55') lassen sich folgende heuristische Argumente anführen: Es ist [3]

$$(1.56)\quad \sigma(\tau)\ (N^T(\tau) + N^T(-\tau)) \geq -\ |\sigma(\tau)|\ \left(\sum_{t=1}^{T-\tau} x_{t+\tau}x'_{t+\tau} + \sum_{t=1}^{T-\tau} x_t x'_t\right) \quad (\tau=1,\ldots,T-1).$$

Zum anderen gilt

$$(1.57)\quad N^T(\tau) + N^T(-\tau) = \sum_{t=1}^{T-\tau} x_{t+\tau}x'_{t+\tau} + \sum_{t=1}^{T-\tau} x_t x'_t - \sum_{t=1}^{T-\tau} (x_{t+\tau}-x_t)(x_{t+\tau}-x_t)'$$

$$(\tau=1,\ldots,T-1).$$

Folgen nun - bei homogener Regression - die Regressoren linearen oder allgemeiner polynomialen Trends, d.h. ist

$x_{it} = q_{im_i} t^{m_i} + \ldots + q_{i1} t + q_{io} = P_{i,m_i}(t)$ mit $m_i \geq 1$, $q_{im_i} \neq 0$ $(i=1,\ldots,n)$,

dann ist wegen $x_{i,t+\tau} = \sum_{j=0}^{m_i} q_{ij}(t+\tau)^j = \sum_{j=0}^{m_i} q_{ij} \sum_{l=0}^{j} \binom{j}{l} t^l \tau^{j-l}$

$x_{i,t+\tau} - x_{it} = \tau\, m_i\, q_{im_i} t^{m_i-1} + P_{m_i-2}(t)$, worin $P_{m_i-2}(t)$

ein Polynom in t vom Grad $\leq m_i-2$ bezeichnet. Daher wird

---

1) Über die mißlichen Folgen auf die Schätzgenauigkeit s. 1.1.2.

2) Für $0 < \tau \leq \tau_o << T$, kann $N^T(\tau)$ - wie auch $N^T(-\tau)$ - (bis auf den Faktor $1/(T-\tau) \doteq 1/T$) als, wenn auch absolutes, (empirisches) Moment der Ordnung τ des x_t-Prozesses angesehen werden. Wesentlich echte Autokovarianzen sind aber in der inhomogenen Regression zu betrachten (vgl (1.64)).

3) Zum Beweis vgl. S. 16.

$$\sum_{t=1}^{T-\tau} (x_{i,t+\tau} - x_{it})(x_{j,t+\tau} - x_{jt}) =$$

$$= \tau^2 m_i m_j \, q_{im_i} \, q_{jm_j} \sum_{t=1}^{T-\tau} t^{\,m_i+m_j-2} + \sum_{t=1}^{T-\tau} P_{m_i+m_j-3}(t)$$

$$= \tau^2 m_i m_j \, q_{im_i} \, q_{jm_j} \, (T-\tau)^{m_i+m_j-1} / (m_i+m_j-1) + O((T-\tau)^{m_i+m_j-2}). \quad 1)$$

Andererseits erhält man

$$\sum_{t=1}^{T-\tau} x_{i,t+\tau} \, x_{j,t+\tau} + \sum_{t=1}^{T-\tau} x_{it} \, x_{jt} = 2 \sum_{t=1}^{T} x_{it} \, x_{jt} - \sum_{t=1}^{\tau} x_{it} \, x_{jt} - \sum_{t=T-\tau+1}^{T} x_{it} \, x_{jt} =$$

$$= 2 \, q_{im_i} \, q_{jm_j} \sum_{t=1}^{T} t^{\,m_i+m_j} + 2 \sum_{t=1}^{T} P_{m_i+m_j-1}(t) - \sum_{t=1}^{\tau} x_{it} \, x_{jt} - \sum_{t=T-\tau+1}^{T} x_{it} \, x_{jt}$$

$$= 2 \, q_{im_i} \, q_{jm_j} \, T^{\,m_i+m_j+1} / (m_i+m_j+1) + 2 \cdot O(T^{\,m_i+m_j}) - \sum_{t=1}^{\tau} x_{it} \, x_{jt} - \sum_{t=T-\tau+1}^{T} x_{it} x_{jt}. \quad 1)$$

Für $0 \leq \tau << T$ (d.h. auch: bei nicht allzu kleinem T) ersieht man aus diesen beiden Formeln, daß in diesem Fall in der Berechnung von $N^T(\tau) + N^T(-\tau)$ $\sum_{t=1}^{T-\tau} (x_{t+\tau} - x_t)(x_{t+\tau} - x_t)'$ gegenüber $\sum_{t=1}^{T-\tau} (x_{t+\tau} x_t' + x_t \, x_{t+\tau}')$ praktisch vernachlässigt und letzterer Ausdruck wiederum praktisch mit $2 X'X$ gleichgesetzt werden kann.[2] $N^T(\tau) + N^T(-\tau)$ ist dann also positiv-definit. Ist dies etwa für $1 \leq \tau \leq \tau_o$ der Fall [3] und sind $\sigma(1), \ldots, \sigma(\tau_o) > 0$, dann ist erst recht $\sum_{\tau=1}^{\tau_o} \sigma(\tau)(N^T(\tau) + N^T(-\tau))$ positiv-definit, während $\sum_{\tau=\tau_o+1}^{T-1} \sigma(\tau)(N^T(\tau) + N^T(-\tau))$ nach (1.56) durch

---

1) nach (4.2.7)

2) Natürlich ist die Argumentation nur für $T \to \infty$ streng.

3) was selbstverständlich $T - \tau_o \geq n$ verlangt.

$$- \sum_{\tau=\tau_o+1}^{T-1} |\sigma(\tau)| \left(\sum_{t=\tau_o+1}^{T} x_t x_t' + \sum_{t=1}^{T-\tau_o} x_t x_t' \right)$$ nach unten abgeschätzt

werden kann. Somit spricht einiges für die Gültigkeit von (1.55') in diesem Fall.

Häufig lassen sich nun (makro-) ökonomische Zeitreihen gut durch einen Polynomtrend approximieren; oft genügt bereits ein linearer Trend zu guter Näherung der Daten. D.h. man kann davon ausgehen, daß der systematische Teil $P_{i,m_i}(t)$ einer Kleinstquadratschätzung $x_{it} = P_{i,m_i}(t) + w_{it}$ $(i = 1,\ldots,n)$ mit einem Polynom $P_{i,m_i}(t)$ - gegebenenfalls geeignet zu bestimmenden Grades ≥ 1 - x_{it} relativ genau - d.h. mit $\sum_{t=1}^{T} w_{it}^2 << \sum_{t=1}^{T} x_{it}^2$ - beschreibt.[1] Aufgrund der obigen Bemerkungen wird man dann von (1.55') ausgehen können.

Dies wird man i.d.R. auch tun können, wenn die Regression ein Absolutglied enthält. Sind nämlich die Daten $z_t' := (x_{2t},\ldots,x_{nt})$ der eigentlichen Regressoren vom angeführten Typ, so ist nach (1.57)

$$N^T(\tau) + N^T(-\tau) = \begin{bmatrix} 2(T-\tau) & \sum_{t=1}^{T-\tau} z_{t+\tau}' + \sum_{t=1}^{T-\tau} z_t' \\ \sum_{t=1}^{T-\tau} z_{t+\tau} + \sum_{t=1}^{T-\tau} z_t & \sum_{t=1}^{T-\tau} z_{t+\tau} z_{t+\tau}' + \sum_{t=1}^{T-\tau} z_t z_t' - \sum_{t=1}^{T-\tau} (z_{t+\tau}-z_t)(z_{t+\tau}-z_t)' \end{bmatrix}$$

$$= 2X'X - \begin{bmatrix} 2\tau & \sum_{t=1}^{\tau} z_t' + \sum_{t=T-\tau+1}^{T} z_t' \\ \sum_{t=1}^{\tau} z_t' + \sum_{t=T-\tau+1}^{T} z_t' & \sum_{t=1}^{\tau} z_t z_t' + \sum_{t=T-\tau+1}^{T} z_t z_t' + \sum_{t=1}^{T-\tau} (z_{t+\tau}-z_t)(z_{t+\tau}-z_t)' \end{bmatrix}.$$

Ähnlich wie oben kann man hieraus (1.55') begründen.

---

1) Ob die Daten tatsächlich von einem - evtl. mit Zufallsstörung versehenen - Trendmodell "produziert" werden, ist in diesem Zusammenhang (bei festem T) unwichtig.

Somit hat man in der Zeitreihenregression - für die trendbehaftete Regressoren und eine positiv autokorrelierte Störvariable typisch sind- häufig davon auszugehen, daß dür die Varianz der Kleinstquadratschätzung gilt:

$$(1.58)\quad V(\hat{\beta}|X) = (X'X)^{-1}\, X'\Sigma X\, (X'X)^{-1} \geq \sigma^2 (X'X)^{-1} \quad .$$

Wenn (1.58) gilt, folgt übrigens des weiteren für die Schätzung (1.7) für σ^2:

$$E(\hat{\sigma}^2|X) = (T-n)^{-1}\, E\,(u'P_X\, u\,|X) = \sigma^2 (T-n)^{-1}\, Sp\; P_X\, \Omega$$

$$= \sigma^2 (T-n)^{-1}\, \{\, T - Sp\; (X'X)^{-1}\, X'\Omega X \,\}\,, \quad \text{d.h.}$$

$$(1.59)\quad E(\hat{\sigma}^2|X) < \sigma^2 \quad .\ ^{1)}$$

Ferner wird die Varianz von $\hat{\beta}$ durch $\hat{\sigma}^2 (X'X)^{-1}$ ebenfalls systematisch unterschätzt:

$$(1.60)\quad E\left[\hat{\sigma}^2 (X'X)^{-1}|X\right] \leq V(\hat{\beta}|X),$$

und zwar, weil sowohl $(X'X)^{-1} \leq (X'X)^{-1}\, X'\Omega\, X\, (X'X)^{-1}$ als auch $E(\hat{\sigma}^2|X) \leq \sigma^2$.

1.2.2.3. Varianz bei Regressoren mit linearem Trend

Um einen Begriff von der Ungenauigkeitswirkung einer positiv autokorrelierten Störvariable zu geben, werde die Varianz der Kleinstquadrat-Schätzung der eigentlichen Regressionskoeffizienten einer inhomogenen Regression, deren eigentliche Regressoren $z_t' = (x_{2t}, \ldots, x_{nt})$ sämtlich mit einem linearen Trend behaftet sind, genauer berechnet.

---

1) $\Omega := \Sigma/\sigma^2$. - Es ist $X'\Omega X = X'X + D$ mit einer positiv-semidefiniten Matrix D und daher $Sp(X'X)^{-1}X'\Omega X = n + Sp(X'X)^{-\frac{1}{2}} D (X'X)^{-\frac{1}{2}} > n$.

Werden Abweichungen vom (empirischen) Mittelwert mit einem "*" bezeichnet und ist Z_* die Matrix mit den Zeilen z_{*t} $(t=1,\ldots,T)$, dann ist

$$\hat{\gamma} = (Z_*' Z_*)^{-1} Z_*' y = \gamma + (Z_*' Z_*)^{-1} Z_*' u$$

die Kleinstquadratschätzung des Vektors γ der eigentlichen Regressionskoeffizienten. Sie hat demnach die Varianz

$$V(\hat{\gamma}|X) = (Z_*' Z_*)^{-1} Z_*' \Sigma Z_* (Z_*' Z_*)^{-1}. \tag{1.61}$$

Da andererseits $V(\hat{\gamma}|X)$ die rechte untere Submatrix von $V(\hat{\beta}|X)$ und $(Z_*' Z_*)^{-1}$ die entsprechende Submatrix von $(X'X)^{-1}$ ist, hat die Gültigkeit von (1.58) zur Folge:

$$Z_*'\Sigma Z_* \geq \sigma^2 Z_*' Z_*. \tag{1.62}$$

Definiert man analog wir früher $N_*^T(\tau) := \sum_{t=1}^{T-\tau} z_{*t+\tau} z_{*t}'$ und $N_*^T(-\tau) := (N_*^T(\tau))'$ $(0 \leq \tau \leq T-1)$, so ist (1.62) gleichbedeutend mit der Positiv-Semidefinitheit von

$$Z_*'\Sigma Z_* - \sigma^2 Z_*' Z_* = \sum_{\tau=1}^{T-1} \sigma(\tau) \quad (N_*^T(\tau) + N_*^T(-\tau)).$$

Mittels der Bezeichnung $\bar{z} := \frac{1}{T} \sum_{t=1}^{T} z_t$, $\bar{z}_{(\tau)1} := \frac{1}{(T-\tau)} \sum_{t=1}^{T-\tau} z_t$, $\bar{z}_{(\tau)2} := \frac{1}{(T-\tau)} \sum_{t=\tau+1}^{T} z_t$ kann man hierin noch die empirischen Autokovarianzmatrizen

$$L_*^T(\tau) := \frac{1}{(T-\tau)} \sum_{t=1}^{T-\tau} (z_{t+\tau} - \bar{z}_{(\tau)2})(z_t - \bar{z}_{(\tau)1})' \quad \{L_*^T(-\tau) := (L_*^T(\tau))';\ \tau=,\ldots,T-1\} \tag{1.63}$$

der Beobachtungsreihen z_t einführen, wodurch man für $\tau=1,\ldots,T-1$ erhält:

$$N_*^T(\tau) + N_*^T(-\tau) = (T-\tau)\left[L_*^T(\tau) + L_*^T(-\tau) + (\bar{z}_{(\tau)2} - \bar{z})(\bar{z}_{(\tau)1} - \bar{z})' + (\bar{z}_{(\tau)1} - \bar{z})(\bar{z}_{(\tau)2} - \bar{z})'\right], \tag{1.64}$$

Diese Formel ermöglicht nun wiederum die folgende erneute Begründung von (1.62): Sind die Beobachtungsvektoren z_t positiv

autokorreliert,[1] so ist für nicht zu kleines T und kleines τ $L_*^T(\tau) + L_*^T(-\tau)$ positiv-definit. Sofern - grob gesprochen - $x_{i\,t}^2 << \sum_{s=1}^{T} x_{i\,s}^2$ $(i=2,\ldots,n)$[2], ist dann auch $N_*^T(\tau) + N_*^T(-\tau)$ positiv-definit; diese Bedingung sorgt aber auch dafür, daß die Summe $\sum_{\tau=1}^{T-1} \sigma(\tau)(N_*^T(\tau) + N_*^T(-\tau))$ von ihren ersten Summanden dominiert wird. Sie ist also selbst positiv-definit.

Gerade im Fall (polynom-) trendbehafteter eigentlicher Regressoren mögen die erwähnten Voraussetzungen gegeben sein, wie nunmehr auch für den einfachsten derartigen Fall, den der linearen Trendbehaftung, verdeutlicht werden soll. Es sei also $x_{1t} = 1$ $(t = 1,\ldots,T)$ und

$$x_{it} = a_i + q_i t + w_{it} \qquad (t = 1,\ldots,T;\ i = 2,\ldots,n).$$

a_i und q_i $(\neq 0)$ mögen darin genauer nach der Kleinstquadrat-

---

1) Als ein Ausdruck der allgemein vorzufindenden positiven Auto- und Reihenkorrelation (makro-)ökonomischer Zeitreihendaten (s. [Schulmann]). - Bei positiver Autokorrelation der x_{it} $(i=2,\ldots,n)$ für kleines τ ist zumindest die Diagonale von $L_*^T(\tau)$ (und $L_*^T(-\tau)$) positiv. Ist dann etwa noch die Autokorrelation der x_{it} stärker ausgeprägt als ihre Reihenkorrelation - m.a.W., dominieren die Diagonalglieder die quadratische Form $c'(L_*^T(\tau) + L_*^T(-\tau))c$ - so ist auch $L_*^T(\tau) + L_*^T(-\tau)$ positiv-definit.

2) Offenbar gilt für die Elemente $\cdot_{ik}$ von $(\bar{z}_{(\tau)2} - \bar{z})(\bar{z}_{(\tau)1} - \bar{z})'$:

$$|\cdot_{ik}| \leq (\max_{t\leq T} z_{it} - \min_{t\leq T} z_{it})(\max_{t\leq T} z_{kt} - \min_{t\leq T} z_{kt}) \quad (i,k=1,\ldots,n-1).$$

Dagegen ist das Element $\cdot_{ik}$ von $L_*^T(\tau)$ eine Summe mit $(T-\tau)$ Summanden.

Methode geschätzte Trendkoeffizienten sein. w_{it} bezeichnet folglich das Trendresiduum und es gilt

$$\sum_{t=1}^{T} w_{it} = 0 , \quad \sum_{t=1}^{T} t\, w_{it} = 0 \quad (i = 2,\ldots,n).$$

Mit den Bezeichnungen $a' := (a_2,\ldots,a_n)$, $q' := (q_2,\ldots,q_n)$, $w_t' := (w_{2t},\ldots,w_{nt})$ $(t = 1,\ldots,T)$ ist $z_t = a + qt + w_t$ und

$$z_{*t} = qt_* + w_t \qquad (t = 1,\ldots,T).$$

Damit wird

$$N_*^T(\tau) = qq' \sum_{t=1}^{T-\tau} (t_*+\tau)t_* + q \sum_{t=1}^{T-\tau} t_* w_t' + \tau q \sum_{t=1}^{T-\tau} w_t' + \sum_{t=1}^{T-\tau} w_{t+\tau}(t_*+\tau)q'$$

$$- \tau \sum_{t=1}^{T-\tau} w_{t+\tau}\, q' + \sum_{t=1}^{T-\tau} w_{t+\tau} w_t'$$

$$= qq' \sum_{t=1}^{T-\tau} t_*^2 - \tau\, qq' \sum_{t=T-\tau+1}^{T} t_* - q \sum_{t=T-\tau+1}^{T} t_* w_t' - \tau q \sum_{t=T-\tau+1}^{T} w_t' - \sum_{t=1}^{\tau} t_* w_t\, q'$$

$$+ \tau \sum_{t=1}^{T} w_t\, q' + \sum_{t=1}^{T-\tau} w_{t+\tau} w_t'. \qquad (0 \le \tau \le T-1)$$

Bei enger Trendbindung der x_{it} (d.h. $w_{it} \doteq 0$), nicht zu kleinem T und $\tau \ll T$, kommt es nur auf den ersten Term an, der die Ordnung $(T-\tau)^3 \doteq T^3$ hat, während etwa

$$\Big|\sum_{t=T-\tau+1}^{T} t_* w_{it}\Big| \text{ durch } \tau \frac{T-1}{2} \max_{t \le T} |w_{it}|, \quad \Big|\sum_{t=T-\tau+1}^{T} w_{it}\Big| \text{ durch } \tau \max_{t \le T} |w_{it}|$$

und $\Big|\sum_{t=T-\tau+1}^{T} t_*\Big|$ durch $\tau \frac{T-1}{2}$ abgeschätzt werden können. Jedenfalls können aber die Terme, in denen w_t's auftreten, vernachlässigt werden. Die ersten Summanden $N_*^T(1)$, $N_*^T(2)$,... dominieren auch $\sum_{\tau=1}^{T-1} \sigma(\tau)(N_*^T(\tau) + N_*^T(-\tau))$, zumal dann, wenn $|\sigma(\tau)|$ mit wachsendem τ rasch abfällt. Bei (zunächst) positiven $\sigma(\tau)$ kann also (1.62) in der Tat als gültig angesehen werden.

Für den autoregressiven Prozeß (1.42) der Störvaiable - $u_t = \varrho u_{t-1} + \varepsilon_t$ $(0<\varrho<1)$ - gewinnt man[1] - unter der nach den obigen Bemerkungen bei enger Trendbindung der Regressoren zulässigen Vernachlässigung der w_{it} $(t=1,\ldots,T;\ i=2,\ldots,n)$ - :

$$\sum_{\tau=1}^{T-1} \sigma(\tau)\,(N_*^T(\tau) + N_*^T(-\tau)) = 2\sigma^2\, qq' \sum_{\tau=1}^{T-1} \varrho^\tau \sum_{t=1}^{T-\tau} (t_*+\tau)\, t_*$$

$$= 2\sigma^2\, qq' \sum_{\tau=1}^{T-1} (T-\tau)\,\varrho^\tau \Big[\frac{1}{T-\tau} \sum_{t=\tau+1}^{T} (t-\bar{t}_{(\tau)2})(t-\bar{t}_{(\tau)1}) + (\bar{t}_{(\tau)2} - \bar{t})(\bar{t}_{(\tau)1} - \bar{t})\Big].$$

Benutzt werden nun weiter die Formeln von 4.2.1. und 4.2.2.; dann ergibt sich

$$\sum_{\tau=1}^{T-1} \sigma(\tau)\,(N_*^T(\tau) + N_*^T(-\tau)) = 2\sigma^2 qq' \sum_{\tau=1}^{T-1} (T-\tau)\,\varrho^\tau \Big[\frac{(T-\tau)^2-1}{12} - \frac{\tau^2}{4}\Big]$$

$$= \frac{\sigma^2 qq'}{6} \sum_{\tau=1}^{T-1} \varrho^\tau \left[(T-\tau)^3 - (T-\tau)(3\tau^2+1)\right]$$

$$= \frac{\sigma^2 qq'}{6} \Big[T(T^2-1) \sum_{\tau=1}^{T-1} \varrho^\tau - (3T^2-1) \sum_{\tau=1}^{T-1} \varrho^\tau \tau + 2 \sum_{\tau=1}^{T-1} \varrho^\tau\, \tau^3\Big]$$

oder

(1.65)
$$\sum_{\tau=1}^{T-1} \sigma(\tau)\,(N_*^T(\tau) + N_*^T(-\tau))$$

$$= \frac{\sigma^2 qq'}{6} \Big[T(T^2-1)\,\varrho\,(1-\varrho^{T-1})(1-\varrho)^{-1}$$

$$- (3T^2-1)\,\varrho\,\{1-\varrho^{T-1}(1+(T-1)(1-\varrho))\}(1-\varrho)^{-2}$$

$$+ 2\varrho\{1-\varrho^{T-1}(1+(T-1)(1-\varrho))^3 - 3T\varrho^T(1-\varrho) + 4\varrho(1-\varrho^T) + \varrho^2(1-\varrho^{T-2})\}(1-\varrho)^{-4}\Big]$$

Dies ist eine positiv-semidefinite Matrix, sobald der Ausdruck in [.] positiv ist. Zur vereinfachten Berechnung desselben kann

---

1) vgl. (1.64)

man ϱ^T und auch $T\varrho^T$ null setzen; dann ist er gleich

$$\varrho(1-\varrho)^{-4}\left[T(T^2-1)(1-\varrho)^3 - (3T^2-1)(1-\varrho)^2 + 2(1+4\varrho+\varrho^2)\right]$$

$$= \varrho(1-\varrho)^{-4}\left[T^2(1-\varrho)^2\{(T-1/T)(1-\varrho) - (3-1/T^2)\} + 2(1+4\varrho+\varrho^2)\right].$$

Diese Größe ist bei $\varrho = .9$ für $T>30$ sicherlich positiv.

Während in diesem Fall der Trendteil von z_t $Z_*'\Sigma Z_*$ wesentlich bestimmt, wird $(Z_*'Z_*)^{-1}$ wesentlich durch die Trendresiduen bestimmt. Dies sei noch kurz dargestellt, um den Kollinearitätseffekt der Trendabhängigkeit zu illustrieren (vgl. 1.1.2.).

Wegen $\sum_{t=1}^{T} w_t = 0 = \sum_{t=1}^{T} t_* w_t$ ist

$$Z_*'Z_* = N_*^T(0) = qq'\sum_{t=1}^{T} t_*^2 + \sum_{t=1}^{T} w_t w_t' = T s_t^2 qq' + \sum_{t=1}^{T} w_t w_t'$$

und somit nach (4.1.5')

$$(1.66)\quad (Z_*'Z_*)^{-1} = \left(\sum_{t=1}^{T} w_t w_t'\right)^{-1} - \frac{T s_t^2}{1+Ts_t^2 q'(\sum_{t=1}^{T} w_t w_t')^{-1} q}\left(\sum_{t=1}^{T} w_t w_t'\right)^{-1} qq'\left(\sum_{t=1}^{T} w_t w_t'\right)^{-1}$$

$$= \frac{1}{T}\left[M_w^{-1} - \frac{1}{1/s_t^2 + q' M_w^{-1} q} M_w^{-1} qq' M_w^{-1}\right]$$

mit $M_w := 1/T \sum_{t=1}^{T} w_t w_t'$ und $s_t^2 := 1/T \sum_{t=1}^{T} (t-\bar{t})^2$.

Das i-te Diagonalelement a_{ii} von $(Z_*'Z_*)^{-1}$ - das im autokorrelationsfreien Modell ja die Varianz $V(\hat{\gamma}|X) = V(\hat{\beta}_{i+1}|X)$ der Schätzung des (i+1)-ten Regressionskoeffizienten $\gamma_i = \beta_{i+1}$ (i = 1,...,n-1) darstellt, hier aber kleiner als $V(\hat{\gamma}_i|X)$ ist - lautet (wenn e_i der i-te Einheitsvektor ist):

$$a_{ii} = \frac{1}{T(1/s_t^2 + q' M_w^{-1} q}\left[1/s_t^2 \; e_i' M^{-1}_w e_i + e_i' M_w^{-1} e_i \; q' M_w^{-1} q - (e_i' M_w^{-1} q)^2\right].$$

Nun ist nach (4.1.7.) $(e_i' M_w^{-1} q)^2 \leq e_i' M_w^{-1} e^i \cdot (M_w^{-1} q)' (M_w^{-1})^{-1} (M_w^{-1} q)$

$$= e_i' M_w^{-1} e_i \cdot q' M_w^{-1} q \; .$$

Ist daher m_w^{ii} das i-te Diagonalelement von M_w^{-1}, so gilt

$$(1.67) \quad a_{ii} \geq \frac{1}{T(1+s_t^2 q' M_w^{-1} q)} \; m_w^{ii} \geq \frac{1}{T s_{wi}^2 (1+s_t^2 q' M_w^{-1} q)} \qquad (i = 1,\ldots,n-1)$$

mit $s_{wi}^2 := 1/T \sum_{t=1}^{T} w_{it}^2$. [1]

1.2.2.4. Bedingungen für die Effizienz der Kleinstquadrat-Schätzung

Von den vielen in der Literatur angebotenen hinreichenden und notwendigen Bedingungen[2] für die Effizienz der Kleinstquadrat-Schätzung in der linearen Schätzklasse bei gegebener positiv-definiter Varianz $E(uu'|X) = \Sigma$ seien hier nur vier wichtige aufgeführt.

Die ersten beiden sind wohlbekannt. Sie sind im R^T formuliert:

Theorem 1.3.[2] Die Kleinstquadrat-Schätzung ist genau dann beste lineare unverzerrte Schätzung, wenn eine der beiden folgenden Bedingungen erfüllt ist:

(i) Es gibt n Eigenvektoren von Σ, aus denen sich jede Spalte von X linear kombinieren läßt.

(ii) $M(\Sigma X) \subset M(X)$.

---

1) s_{wi}^2 ist das i-te Diagonalelement von M_w; daher ist $m_w^{ii} \geq 1/s_{wi}^2$ Wie man beispielsweise (4.1.1) entnimmt).

2) Vgl. insbesondere [Rao], Sect. 8.2 .

Beweis[1]: Gleichheit der Varianz von Kleinstquadrat-und Aitken-Schätzung ist (s. 1.2.1.1.) gleichbedeutend mit $(X'X)^{-1}X' = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}$ oder

$$\Sigma X = X(X'\Sigma^{-1}X)^{-1}X'X. \tag{1.68}$$

Ist Λ Diagonalmatrix mit den (positiven) Eigenwerten des Problems $X'\Sigma^{-1}X - \lambda X'X$ auf der Diagonalen und V mit $V'X'XV = I$ die $(n \times n)$-Matrix mit den zugehörigen Eigenvektoren als Spalten, so folgt aus (1.68) $\Sigma XV = XV\,V^{-1}(X'\Sigma^{-1}X)^{-1}V^{-1}V'X'XV = XV\Lambda^{-1}$. Dies bedeutet, daß $W = XV$ eine Matrix mit n Eigenvektoren von Σ als Spalten ist, aus denen sich die Spalten von X gemäß $X = WV^{-1}$ linear kombinieren lassen. Ist umgekehrt W $(T \times n)$-Matrix vom Rang n mit $X = WH$ und $\Sigma W = W\mu$ (μ diagonal), dann gilt $\Sigma X = \Sigma WH = W\mu H = XH^{-1}\mu H$ oder $\Sigma X = XB$, woraus $X = \Sigma^{-1}XB$, $B = (X'\Sigma^{-1}X)^{-1}X'X$, also (1.68) folgt. - Damit ist der Beweis sowohl bzgl. Bedingung (i) als auch bzgl. Bedingung (ii) vollendet, da letztere ja nichts anderes bedeutet, als daß die Spalten von ΣX sich aus den Spalten von X linear kombinieren lassen.

Weniger bekannt ist eine im R^n formulierte Bedingung, die im folgenden hergeleitet werden soll. - Dazu wird zunächst der Begriff des Spektrums von X'X bzgl. Σ eingeführt. Sind $\mu_1,\ldots,\mu_q$ die verschiedenen, möglicherweise mehrfachen Eigenwerte von Σ und enthalten $V_1,\ldots,V_q$ mit $V_i'V_i = I$, $V_i'V_j = 0$ $(i \neq j,\ i,j=1,\ldots,q)$ die zugehörigen Eigenvektoren als Spalten, dann ist V_iV_i' eindeutig bestimmt $(i=1,\ldots,q)$ und es gilt:

$$I = \sum_{i=1}^{q} V_iV_i' \;;\quad \Sigma = \sum_{i=1}^{q} \mu_iV_iV_i' \;;\quad \Sigma^{-1} = \sum_{i=1}^{q} \frac{1}{\mu_i} V_iV_i' \;. \tag{1.69}$$

Folglich ist mit $E_i := X'V_iV_i'X$ $(i=1,\ldots,q)$

---

1) In Anlehnung an [Anderson], S.561

(1.70) $X'X = \sum_{i=1}^{q} E_i \; ; \quad X'\Sigma X = \sum_{i=1}^{q} \mu_i E_i \; ; \quad X'\Sigma^{-1}X = \sum_{i=1}^{q} \frac{1}{\mu_i} E_i \; .$

Das System S der von Null verschiedenen E_i heißt Spektrum von $X'X$ bzgl. Σ.

S besitzt nun eine eindeutig bestimmte Maximalzerlegung $S = S_1 + \ldots + S_m$ derart, daß für $E_i \in S_k$, $E_j \in S_l$ bei $k \neq l$ stets $E_i(X'X)^{-1}E_j = O$ gilt (k,l=1,...,m) und keine der nichtleeren Mengen S_k (k=1,...,m) sich weiter zerlegen läßt, wenn diese Eigenschaft insgesamt erhalten bleiben soll. - Daß eine Maximalzerlegung existiert, sieht man wie folgt: Zu einem $E_i \in S$, etwa zu E_1, definiere man $\vartheta_1 := \{E_k \in S \mid E_k(X'X)^{-1}E_1 \neq O\}$. Dann ist $\vartheta_1 \neq O$ (nämlich $E_1 \in \vartheta_1$). Ist ϑ_1 bereits definiert, dann erkläre man $\vartheta_{1+1} := \{E_k \in S - (\vartheta_1 \vee \ldots \vee \vartheta_1) \mid E_k(X'X)^{-1}E_j \neq O$ für ein $E_j \in \vartheta_1\}$, sofern diese Menge nicht leer ist. Nach Konstruktion gilt für $E_k \in S - (\vartheta_1 \vee \ldots \vee \vartheta_1)$ $E_k(X'X)^{-1}E_j = O$ für jedes $E_j \in \vartheta_1, \ldots, \vartheta_{1-1}$. Da die Mengen ϑ nicht leer und disjunkt sind, bricht das Verfahren nach $r \leq q \leq T$ Schritten ab. Man setzt dann $S_1 := \vartheta_1 + \ldots + \vartheta_r$ und $S_2 := S - S_1$. Ist S_2 leer, dann hat S offenbar nur die triviale Maximalzerlegung S = S. Ist dagegen S_2 nicht leer, so liegt mit $S = S_1 + S_2$ eine Zerlegung vor, die $E_k(X'X)^{-1}E_j = O$ für jedes $E_k \in S_1$, $E_j \in S_2$ erfüllt. Hat man nun bereits eine Zerlegung $S = S_1 + \ldots + S_p$ mit der entsprechenden Eigenschaft für alle Paare S_l, S_h ($l \neq h$, l,h=1,...,p) vorliegen, so führt jede mögliche Zerlegung einer der Teilmengen S_l (l=1,...,p) nach dem obigen Muster zu einer verfeinerten Zerlegung von S selbst (deren Summanden also wiederum die angeführte Eigenschaft besitzen). Schließlich gelangt man, wenn man auf diese Weise verfährt, (nach spätestens q Schritten) zu einer Maximalzerlegung von S. Wären nun $S = S_1 + \ldots S_m$ und $S = S_1' + \ldots S_r'$ zwei verschiedene Maximalzerlegungen, dann gäbe es eine Menge S_i und eine Menge S_l' mit $S_i \cap S_l' \neq \emptyset$ und $S_i \cap (S - S_l') \neq \emptyset$. Da aber für $E_k \in S_l'$ und $E_j \in S - S_l'$ $E_k(X'X)^{-1}E_j = O$ gilt, wäre wegen $S_i = (S_i \cap S_l') + (S_i \cap (S - S_l'))$ die erste Zerlegung nicht maximal. Die Maximalzerlegung ist also eindeutig. -

Ihre Elemente $S_1,\ldots,S_m$ heißen Elemente des Spektrums von $X'X$ bzgl. Σ.[1)]

Nach diesen Vorbereitungen nun zur eigentlichen Aussage. Mit n-Vektoren c und $d(\neq 0)$ und der Bezeichnung $W_i := X'V_i$ $(i=1,\ldots,q)$ wird

$$c'X'Xd = \sum_{i=1}^{q} c'E_i d = \sum_{i=1}^{q} c'W_i W_i' d = \sum_{i=1}^{q} \left(\frac{1}{\sqrt{\mu_i}} W_i' c\right)' \left(\sqrt{\mu_i}\, W_i' d\right).$$

Mittels der Schwarzschen Ungleichung folgt daraus

$$(1.71)\qquad (c'X'Xd)^2 \le \sum_{i=1}^{q} \left(\frac{1}{\sqrt{\mu_i}} W_i' c\right)' \left(\frac{1}{\sqrt{\mu_i}} W_i' c\right) \sum_{i=1}^{q} \left(\sqrt{\mu_i}\, W_i' d\right)' \left(\sqrt{\mu_i}\, W_i' d\right)$$

$$= c' \sum_{i=1}^{q} \frac{1}{\mu_i} W_i W_i' c \cdot d' \sum_{i=1}^{q} \mu_i W_i W_i' d = c'X'\Sigma^{-1}Xc \cdot d'X'\Sigma Xd\ .$$

Das Gleichheitszeichen gilt hierin (da wegen $\sum_{i=1}^{q} W_i W_i' d = X'Xd$ nicht alle $W_i'd$ null sein können) genau dann, wenn

$$(1.72)\qquad \frac{1}{\sqrt{\mu_i}} W_i' c = k' \sqrt{\mu_i}\, W_i' d \qquad (k' \neq 0;\ \ i = 1,\ldots,q)$$

ist, worin k' von i nicht abhängt.

Aus (1.71) folgt unter Benutzung von (4.1.7)

$$d'X'\Sigma Xd \ge \max_{c \neq 0} \frac{(c'X'Xd)^2}{c'X'\Sigma^{-1}Xc} = d'X'X(X'\Sigma^{-1}X)^{-1}X'Xd\,;$$

das Maximum wird genau für die c mit $c = k''\,(X'\Sigma^{-1}X)^{-1}X'Xd$ $(k'' \neq 0)$ angenommen.

Damit ergibt sich als hinreichende und notwendige Bedingung für die effiziente Schätzung von $\alpha = d'\beta$ nach der Kleinstquadrat-Methode

---

1) Zur Definition s. [Watson], S.1691.- Tatsächlich muß nicht nur $m \le q \le T$, sondern sogar $m \le n$ gelten, worauf hier aber nicht eingegangen werden soll.

(1.73) $W_i'(X'\Sigma^{-1}X)^{-1}X'Xd = k\,\mu_i\,W_i'd \qquad (k\neq 0 \text{ unabhängig von } i;\ i=1,\ldots,q).$

Multipliziert man diese Gleichung mit $1/\mu_i\,W_i$ und summiert über i, dann folgt $X'Xd = kX'Xd$. k hängt also von d gar nicht ab, sondern ist vielmehr gleich 1, und (1.73) ist gleichwertig mit

(1.73') $W_i'(X'\Sigma^{-1}X)^{-1}X'Xd = \mu_i\,W_i'd \qquad (i = 1,\ldots,q).$

Effiziente Schätzung des gesamten Vektors β ist daher gleichbedeutend mit $W_i'(X'\Sigma^{-1}X)^{-1}X'X = \mu_i\,W_i' \qquad (i = 1,\ldots,q)$ oder

(1.74) $(X'\Sigma^{-1}X)^{-1}W_i = \mu_i\,(X'X)^{-1}W_i \qquad (i = 1,\ldots,q)$

Ist nun die Kleinstquadrat-Schätzung effizient, dann folgt nach (1.74) (nach Multiplikation mit W_i'): Für $E_i \in S$ (d.h. $E_i \neq 0$) ist μ_i Eigenwert des Problems $(X'\Sigma^{-1}X)^{-1} - \lambda\,(X'X)^{-1}$ und die (von null verschiedenen) Spalten von E_i sind zugehörige Eigenvektoren. Sind deshalb E_i, $E_j \in S$, so folgt aus $i\neq j$ wegen $\mu_i \neq \mu_j$ $E_i(X'X)^{-1}E_j = 0$. - Ist umgekehrt bei $i\neq j$ für $E_i, E_j \in S$ stets $E_i(X'X)^{-1}E_j = 0$, dann ist für jedes $E_i \in S$ nach (1.70) $E_i = E_i(X'X)^{-1}X'X = \sum_S E_i(X'X)^{-1}E_j = E_i(X'X)^{-1}E_i$ und $X'\Sigma^{-1}X(X'X)^{-1}E_i = \sum_S \frac{1}{\mu_j} E_j(X'X)^{-1}E_i = \frac{1}{\mu_i}E_i$. D.h. es ist $X'X(X'\Sigma^{-1}X)^{-1}E_i = \mu_i E_i$ und nach Summation über S $X'X(X'\Sigma^{-1}X)^{-1}X'X = X'\Sigma X$; die Kleinstquadrat-Schätzung ist also effizient.

Theorem 1.4. (Watson). Die Kleinstquadrat-Schätzung ist genau dann beste lineare unverzerrte Schätzung, wenn die folgende Bedingung gilt:

(i) Für $E_i, E_j \neq 0$, $i\neq j$ $(i,j=1,\ldots,q)$ ist stets $E_i(X'X)^{-1}E_j = 0$. Sie läßt sich auch formulieren als:

(i') Jedes Element des Spektrums von X'X bzgl. Σ ist eine einelementige Menge.

(ii') Die Eigenwerte von Σ sind konstant auf den Elementen des Spektrums von X'X bzgl. Σ.

Enthält eine Element S_1 des Spektrums S von X'X bzgl. Σ die verschiedenen Matrizen $E_{k1}, E_{k2}, \ldots$, dann betrachtet man die Eigenwerte $\mu_{k1}, \mu_{k2}, \ldots$ als S_1 zugeordnet. (ii') besagt, daß jedem S_1 nur ein Eigenwert zugeordnet ist; es handelt sich also um eine bloße Beschreibung des Sachverhalts (i').

Dieses Theorem wurde in der Formulierung (ii') von Watson[1] als Gegenstück in der endlichen Stichprobe eines (für $T \to \infty$) asymptotisch bestehenden Sachverhalts (des → Theorems 2.4. von Grenander und Rosenblatt nämlich) hergeleitet.[2] Der obige Beweis ist allerdings anders als Watson's Herleitung.

Korollar 1.1. (Anderson). Die Kleinstquadrat-Schätzung ist genau dann beste lineare unverzerrte Schätzung, wenn gilt:

(i) $\sum_S r(E_i) = n.$

Dies läßt sich auch formulieren als:

(i') Dem Spektrum S von X'X bzgl. Σ sind $p \leq n$ Eigenwerte von Σ zugeordnet und die Summe der Ränge der zugehörigen Matrizen E_i ist n.

$r(E_i)$ bezeichnet den Rang E_i.
Die Formulierung (i') ist angelehnt an ein Korollar von Anderson zu dem erwähnten Theorem von Grenander und Rosenblatt, während die Formulierung (i) mit einem Theorem Anderson's identisch ist [Anderson], S.580 bzw. 563); die Herleitung ist allerdings anders gestaltet.

Beweis: Es gelte Bedingung (i). Bezeichnet man
$A_i := (X'X)^{-1/2} E_i (X'X)^{-1/2}$ $(i=1,\ldots,q)$, so gilt wegen

---

1) [Watson], S.1691

2) Um die Analogie noch besser herauszustellen, sollte man eher - wie Watson dies auch tut - das Spektrum S und die Elemente des Spektrums S in den Indizes i definieren, nicht, wie oben geschehen, in den Matrizen E_i.

$\sum_S E_i = X'X \quad \sum_S A_i = I$ und nach Voraussetzung $\sum_S r(A_i) = n.$

Für (n x n)-Matrizen A,B gilt weiterhin stets - wenn 'dim' die Dimension bezeichnet -

$$r(A+B) \leq \dim\{M(A) + M(B)\} = \dim\{M(A)\} + \dim\{M(B)\} - \dim\{M(A) \cap M(B)\}$$

$$= r(A) + r(B) - \dim\{M(A) \cap M(B)\} \leq r(A) + r(B),$$

so daß aus $r(A+B) = r(A) + r(B)$ $\quad M(A) \cap M(B) = O$ folgt.

Nun ist $n = r(I) \leq r(A_i) + r(I-A_i) = r(A_i) + r(\sum_{S-\{E_i\}} A_j) \leq r(A_i) + \sum_{S-\{E_i\}} r(A_j) = n,$

also $M(A_i) \cap M(I-A_i) = O$ und somit für Spalte $a_i^{(k)}$ $(k=1,\ldots,n)$

von A_i wegen $M(A_i) \ni a_i^{(k)} - A_i a_i^{(k)} = (I-A_i)a_i^{(k)} \in M(I-A_i) : A_i a_i^{(k)} = a_i^{(k)}$;

d.h. es ist $A_i^2 = A_i$ für jedes E_i aus S. Weiter folgt

$$A_i = A_i I A_i = A_i^3 + \sum_{S-\{E_i\}} A_i A_j A_i = A_i + \sum_{S-\{E_i\}} A_i A_j A_i ,$$

also $\sum_{S-\{E_i\}} A_i A_j A_i = O$. Da aber A_i symmetrisch ist und die A_j positiv-semidefinit symmetrisch sind, hat dies $A_i A_j = O$ und damit $E_i (X'X)^{-1} E_j = O$ $(E_j \in S-\{E_i\})$ zur Folge. - Gilt umgekehrt $E_i (X'X)^{-1} E_j = O$, also $A_i A_j = O$ für alle $E_i, E_j \in S$ mit $i \neq j$, dann ist, wie man durch Induktion der obigen Formel leicht zeigt,

$$\sum_S r(E_i) = \sum_S r(A_i) = \sum_S \dim\{M(A_i)\} = \dim\{\sum_S M(A_i)\}.$$

Da aber $\sum_S A_i = I$ ist, also $\sum_S a_i^{(k)} = e_k$ (der k-te Einheitsvektor; $k=1,\ldots,n$), ist $\dim\{\sum_S M(A_i)\} = n.$ -

1.2.2.5. Eine Untergrenze der Varianz

Theorem 1.5. Folgt die Störvariable u_t (bei gegebenem X[1]) einem stationären Prozeß mit positiver stetiger Spektraldichte $f(\lambda)$ $(-\pi \leq \lambda \leq \pi)$, dann gilt

---

1) s. dazu Fn. 2, S.11

(1.75) $$V(\hat{\beta}|X) = 2\pi(X'X)^{-1}\int_{-\pi}^{\pi} f(\lambda)\, d\mathcal{G}_T(\lambda)\,(X'X)^{-1} \geq 2\pi\left(\int_{-\pi}^{\pi}\frac{1}{f(\lambda)}\, d\mathcal{G}_T(\lambda)\right)^{-1}.$$

Beweis: $\mathcal{G}_T(\lambda)$ ist die in 1.2.1.2. eingeführte Spektralverteilungsfunktion der Folge (1.19) . Sie hat gemäß (4.5.2) eine Dichte $g_T(\lambda)$. Daß $V(\hat{\beta}|X)$ die angegebene Gestalt hat, ist (1.49') zu entnehmen. Für (reelle) n-Vektoren $c,d(\neq 0)$ ist, da die positiv-semidefinite hermitesche Matrix $g_T(\lambda)$ eine Zerlegung $g_T(\lambda) = m^*(\lambda)\, m(\lambda)$ hat $(-\pi\leq\lambda\leq\pi)$,

$$c'X'Xd = c'\int_{-\pi}^{\pi} d\,\mathcal{G}_T(\lambda)d = c'\int_{-\pi}^{\pi} g_T(\lambda)\, d\lambda\, d = \int_{-\pi}^{\pi}\left(\frac{1}{\sqrt{f(\lambda)}}\, m(\lambda)c\right)^*\left(\sqrt{f(\lambda)}\, m(\lambda)d\right)d\lambda.$$

Mittels der Schwarzschen Ungleichung folgt daraus

(1.76) $$(c'X'Xd)^2 \leq \int_{-\pi}^{\pi}\left(\frac{1}{\sqrt{f(\lambda)}}\, m(\lambda)c\right)^*\left(\frac{1}{\sqrt{f(\lambda)}}\, m(\lambda)c\right)d\lambda \cdot \int_{-\pi}^{\pi}\left(\sqrt{f(\lambda)}\, m(\lambda)d\right)^*\left(\sqrt{f(\lambda)}\, m(\lambda)d\right)\, d\lambda$$

$$= c'\int_{-\pi}^{\pi}\frac{1}{f(\lambda)}\, d\mathcal{G}_T(\lambda)c \cdot d'\int_{-\pi}^{\pi} f(\lambda)\, d\mathcal{G}_T(\lambda)\, d.$$

Die Anwendung von (4.1.7) zeigt sodann, daß für jedes d

(1.77) $$d'\int_{-\pi}^{\pi} f(\lambda)d\mathcal{G}_T(\lambda)d \geq \max_{c\neq 0}\frac{(c'X'Xd)^2}{c'\int_{-\pi}^{\pi}\frac{1}{f(\lambda)}d\mathcal{G}_T(\lambda)c} = d'X'X\left(\int_{-\pi}^{\pi}\frac{1}{f(\lambda)}d\mathcal{G}_T(\lambda)\right)^{-1}X'Xd$$

gilt. Daraus folgt die Behauptung.-

Bevor die hinreichende und notwendige Bedingung für die Gültigkeit des Gleichheitszeichens in (1.75) abgeleitet wird, werde der Begriff des Spektrums S bzgl. einer Spektralverteilungsfunktion $\mathcal{F}(\lambda)$ [1] eingeführt. Man versteht darunter die Menge aller Wachstumsstellen von $\mathcal{F}(\lambda)$, d.i. die Menge aller λ, für die es zu jedem $\varepsilon > 0$ ein Paar $0 < h, h' < \varepsilon$ gibt, so daß $\mathcal{F}(\lambda+h) - \mathcal{F}(\lambda-h')$ positiv-semidefinit, aber nicht null ist. Da für $h,h' > 0$ $\mathcal{F}(\lambda+h) - \mathcal{F}(\lambda-h')$ stets positiv-semidefinit hermitesch ist, folgt aus

---

1) Vgl. S.10 f.

$F_{jj}(\lambda+h) = F_{jj}(\lambda-h')$ $(j=1,\ldots,n)$ bereits $\mathfrak{F}(\lambda+h) = \mathfrak{F}(\lambda-h')$; es ist also $S = \bigcup_{j=1}^{n} S_j$, wenn S_j des Spektrum des Diagonalelements $F_{jj}(\lambda)$ ist $(j=1,\ldots,n)$. Ist λ keine Wachstumsstelle von $\mathfrak{F}$, dann gibt es (für jedes F_{jj}, $j=1,\ldots,n$, und damit überhaupt) ein Intervall maximaler Länge um λ, in dem $\mathfrak{F}$ konstant ist. Daraus folgt, daß S eine Borelmenge ist und daß es bei Integration bzgl. $\mathfrak{F}$ über die $\mathfrak{F}$-meßbare Menge A genügt, die Integration über $A \cap S$ auszuführen. Offenbar ist $S \subset [-\pi,\pi]$ und enthält jede Unstetigkeitsstelle von .

Ist $\mathfrak{F}(\pi) = M$ positiv-definit, dann läßt sich zeigen, daß es eine (bis auf allfällige Ergänzung durch $\mathfrak{F}$-Nullmengen) eindeutige Zerlegung $S = S_1 + \ldots + S_m$ des Spektrums S in $\mathfrak{F}$-meßbare Mengen[1] $S_1,\ldots,S_m$ gibt derart, daß

(i) $\int_{S_k} d\mathfrak{F} \neq 0$ $(k=1,\ldots,m)$

(ii) $\int_{S_k} d\mathfrak{F}\, M^{-1} \int_{S_l} d\mathfrak{F} = 0$ für $k \neq l$ $(k,l=1,\ldots,m)$

(iii) keine der Mengen S_k $(k=1,\ldots,m)$ sich weiter zerlegen läßt; d.h. bei $S_k = S_{k1} + S_{k2}$ mit $\mathfrak{F}$-meßbaren Mengen S_{k1}, S_{k2}, für die $\int_{S_{k1}} d\mathfrak{F}$ und $\int_{S_{k2}} d\mathfrak{F}$ von null verschieden sind, gilt $\int_{S_{k1}} d\mathfrak{F}\, M^{-1} \int_{S_{k2}} d\mathfrak{F} \neq 0$.

Die Mengen $S_1,\ldots,S_m$ dieser Zerlegung heißen Elemente des Spektrums S bzgl. $\mathfrak{F}(\lambda)$. - [2]

Es sei $S = S_1 + \ldots + S_m$ eine Zerlegung mit den Eigenschaften (i), (ii) und eine der Mengen S_k lasse sich weiter zerlegen: $S_k = S_{k1} + S_{k2}$ $(S_{k2} := S_k - S_{k1})$ mit $\int_{S_{k1}} d\mathfrak{F} \neq 0$, $\int_{S_{k2}} d\mathfrak{F} \neq 0$ und

---

1) Da $\mathfrak{F}(\lambda)$ (nämlich für jedes $c \neq 0$ $F_c(\lambda) := c^* \mathfrak{F}(\lambda) c$) ein Intervallmaß definiert, sind alle $\mathfrak{F}$-meßbaren Mengen (bis auf jeweilige Ergänzung durch eine $\mathfrak{F}$-Nullmenge) Borelmengen.

2) Zur Definition vgl. [Grenander und Rosenblatt], S.241.

$\int_{S_{k2}} d\mathcal{F}\, M^{-1} \int_{S_{k2}} d\mathcal{F} = 0$. Dann ist auch $S = S_{k1} + S_{k2} + \sum_{l\neq k,1}^{m} S_k$ eine - und zwar feinere - Zerlegung von S mit den Eigenschaften (i) und (ii).

Denn für $l \neq k$ folgt aus $\int_{S_k} d\mathcal{F} = \int_{S_{k1}} d\mathcal{F} + \int_{S_{k2}} d\mathcal{F}$ und

$\int_{S_l} d\mathcal{F}\, M^{-1} \int_{S_k} d\mathcal{F} = 0$: $\int_{S_l} d\mathcal{F}\, M^{-1} \int_{S_{k1}} d\mathcal{F} = \int_{S_l} d\mathcal{F}\, M^{-1} \int_{S_{k2}} d\mathcal{F} = 0$.

Erklärt man nämlich $A_l := M^{-1/2} \int_{S_l} d\mathcal{F}\, M^{-1/2}$ $(l\neq k;\ l = 1,\dots,m)$

und analog A_k, A_{k1}, A_{k2}, so sind diese Matrizen positiv-semidefinit hermitesch[1] und erfüllen $A_k + \sum_{l\neq k} A_l = I = A_{k1} + A_{k2} + \sum_{l\neq k} A_l$ nebst $A_l A_k = A_l A_s = 0$ $(l,s \neq k)$. Also gilt für $l \neq k$

$A_l = A_k A_l + \sum_{s\neq k,l} A_s A_l + A_l^2 = A_l^2$ und daher

$A_l = A_l I A_l = A_l A_{k1} A_l + A_l A_{k2} A_l + \sum_{s\neq k,l} A_l A_s A_l + A_l^3 = A_l A_{k1} A_l + A_l A_{k2} A_l + A_l$,

woraus $A_l A_{k1} = A_l A_{k2} = 0$ und somit die Behauptung folgt. -

Genau wie im Beweis von Korollar 1.1. folgt weiterhin, wenn $S = S_1 + \dots + S_m$ eine Zerlegung mit den Eigenschaften (i) und (ii) ist:

(1.78) $\sum_{l=1}^{m} r\left(\int_{S_l} d\mathcal{F}\right) = \sum_{l=1}^{m} r\,(A_l) = n.$

Es muß also $m \leq n$ sein. - Diese Betrachtung zeigt, daß es eine Maximalzerlegung von S mit (i) und (ii) gibt, d.i. eine Zerlegung, die daneben noch (iii) erfüllt. Gäbe es nun zwei verschiedene Maximalzerlegungen $S = S_1 + \dots + S_m$ und $S = S_1' + \dots + S_r'$ mit (i) und (ii), dann müßte für eine Menge S_k und eine Menge S_l' $S_{k1} := S_k \cap S_l' \neq \emptyset$, $S_{k2} := S_k \cap (S - S_l') \neq \emptyset$ und auch $\int_{S_{k1}} d\mathcal{F} \neq 0$, $\int_{S_{k2}} d\mathcal{F} \neq 0$ sein. Da $S_k = S_{k1} + S_{k2}$ ist, würde daraus

---

1) Da für $\mathcal{F}$-meßbare B $\left(\int_B d\mathcal{F}\right)^* = \int_B d\mathcal{F}^* = \int_B d\mathcal{F}$ und

$c^* \int_B d\mathcal{F}\, c = \int_B d\,(c^* \mathcal{F} c) \geq 0$ ist; insbesondere gilt dies für $M = \int_{-\pi}^{\pi} d\mathcal{F}$.

wie oben $\int_{S_i} d\mathcal{F}\, M^{-1} \int_{S_{k1}} d\mathcal{F} = \int_{S_i} d\mathcal{F}\, M^{-1} \int_{S_{k2}} d\mathcal{F} = 0$ für jedes $i \neq k$

folgen $(i = 1,\ldots,m)$.

Ferner wäre auch

$$\int_{S_{k1}} d\mathcal{F}\, M^{-1} \int_{S_{k2}} d\mathcal{F} = \int_{S_k \cap S_1'} d\mathcal{F}\, M^{-1} \int \sum_{j\neq 1} d\mathcal{F}_{S_k \cap S_j'} = \int_{S_k \cap S_1'} d\mathcal{F}\, M^{-1} \sum_{j\neq 1} \int_{S_k \cap S_j'} d\mathcal{F}$$

$$= \sum_{j\neq 1} \int_{S_1' \cap S_k} d\mathcal{F}\, M^{-1} \int_{S_j' \cap S_k} d\mathcal{F} = 0 ,$$

da (nach demselben Argument wie oben) wegen $S_j' \cap S_k \subset S_j'$ und $\int_{S_1'} d\mathcal{F}\, M^{-1} \int_{S_j'} d\mathcal{F} = 0$ zunächst $\int_{S_1'} d\mathcal{F}\, M^{-1} \int_{S_j' \cap S_k} d\mathcal{F} = 0$ und dann auch

wegen $S_1' \cap S_k \subset S_1'$ $\int_{S_1' \cap S_k} d\mathcal{F}\, M^{-1} \int_{S_j' \cap S_k} d\mathcal{F} = 0$ $(j \neq 1;\ j = 1,\ldots,r)$

folgen würde. Die erste Zerlegung wäre somit nicht maximal.

Die Einführung dieser Begriffe ermöglicht folgende Aussage:

Theorem 1.6: Unter der Voraussetzung von Theorem 1.1.5. gilt in der Ungleichung (1.75) das Gleichheitszeichen genau dann, wenn $f(\lambda)$ konstant auf den Elementen des Spektrums S bzgl. $\mathcal{G}_T(\lambda)$ ist.

Beweis: Gültigkeit des Gleichheitszeichens in (1.75) ist gleichbedeutend mit $d' \int_{-\pi}^{\pi} f(\lambda)\, d\mathcal{G}_T(\lambda)\, d = d'X'X\left(\int_{-\pi}^{\pi} \frac{1}{f(\lambda)}\, d\,\mathcal{G}_T(\lambda) \right)^{-1} X'Xd$

für jedes d.

Für festes $d \neq 0$ ist diese Gleichheit - aufgrund von (1.77) und (4.1.7) - äquivalent damit, daß für jedes $c = k''\left(\int_{-\pi}^{\pi} \frac{1}{f(\lambda)}\, d\,\mathcal{G}_T(\lambda) \right)^{-1} X'X\, d$ $(k'' \neq 0)$ in (1.76) das Gleichheitszeichen gilt, d.h. äquivalent mit

$$k'' \frac{1}{\sqrt{f(\lambda)}} m(\lambda) \left(\int_{-\pi}^{\pi} \frac{1}{f} d\mathfrak{g}_T \right)^{-1} X'Xd = k' \sqrt{f(\lambda)}\, m(\lambda) d \qquad (k \neq 0) \quad \text{oder}$$

$$m(\lambda) \left(\int_{-\pi}^{\pi} \frac{1}{f} d\mathfrak{g}_T \right)^{-1} X'Xd = k\, f(\lambda)\, m(\lambda)\, d \qquad (k \neq 0) \tag{1.79}$$

für jedes λ in $[-\pi,\pi] - N_d$, worin N_d eine Menge vom Lebesgue-Maß Null ist; k hängt von λ nicht ab.

Multiplikation von (1.79) mit $\frac{1}{f(\lambda)} m^*(\lambda)$ und anschließende Integration über $[-\pi,\pi]$ zeigt, daß k auch von d nicht abhängt, sondern vielmehr gleich 1 ist.

Es gelte nun in (1.75) das Gleichheitszeichen. Dann ist (1.79) für jeden Einheitsvektor $d = e_i$ $(i=1,\ldots,n)$ erfüllt und $\bigcup_{i=1}^{n} N_{e_i}$ hat das Lebesgue-Maß Null. Spaltenweises Zusammenfassen dieser Gleichungen und Multiplikation mit $m^*(\lambda)$ zeigt, daß

$$\mathfrak{g}_T(\lambda) \left(\int_{-\pi}^{\pi} \frac{1}{f} d\mathfrak{g}_T \right)^{-1} X'X = f(\lambda) \mathfrak{g}_T(\lambda) \;, \quad \text{d.h. (wegen } \mathfrak{g}_T^*(\lambda) = \mathfrak{g}_T(\lambda))$$

$$\left(\int_{-\pi}^{\pi} \frac{1}{f} d\mathfrak{g}_T \right)^{-1} \mathfrak{g}_T(\lambda) = f(\lambda)(X'X)^{-1} \mathfrak{g}_T(\lambda) \tag{1.80}$$

für (mit evtl. Ausnahme einer Menge N vom Lebesgue-Maß Null) alle λ ist $(-\pi \leq \lambda \leq \pi)$. Hieraus folgt weiter: Ist $\mathfrak{g}_T(\lambda) \neq 0$, dann ist $f(\lambda)$ Eigenwert des Problems $\left(\int \frac{1}{f} d\mathfrak{g}_T \right)^{-1} - \mu (X'X)^{-1}$. Ist $N_1 = \{-\pi \leq \lambda \leq \pi \mid \mathfrak{g}_T(\lambda) = 0\}$, so nimmt $f(\lambda)$ in $[-\pi,\pi] - (N_1 \cup N)$ also nur $p' \leq n$ verschiedene Werte $c_1,\ldots,c_{p'}$ an. Sind S_k die Mengen[1] $\{\lambda \in [-\pi,\pi] - (N_1 \cup N) \mid f(\lambda) = c_k\}$ $(k = 1,\ldots,p')$, so sei etwa $\int_{S_k} d\mathfrak{g}_T \neq 0$ für $k = 1,\ldots,p$ und $\int_{S_k} d\mathfrak{g}_T = 0$ für $k = p+1,\ldots,p'$.

$N_2 := S_{p+1} + \ldots + S_{p'}$, sowie auch N_1 und N sind $\mathfrak{g}_T$-Nullmengen. $[-\pi,\pi] - (N_2 \cup N_1 \cup N)$ kann daher mit dem Spektrum S bzgl. $\mathfrak{g}_T(\lambda)$ identifiziert werden und $S_1,\ldots,S_p$ definieren eine Zerlegung von S. Aus

$$\left(\int_{-\pi}^{\pi} \frac{1}{f} d\mathfrak{g}_T \right)^{-1} \mathfrak{g}_T(\lambda) = c_k \, (X'X)^{-1} \, \mathfrak{g}_T(\lambda) \qquad (\lambda \in S_k) \qquad (k = 1,\ldots,p)$$

folgt aber durch Integration über S_k:

$$\left(\int_{-\pi}^{\pi} \frac{1}{f} d\mathfrak{g}_T \right)^{-1} \int_{S_k} d\mathfrak{g}_T = c_k \, (X'X)^{-1} \int_{S_k} d\mathfrak{g}_T \qquad (k = 1,\ldots,p).$$

---

1) S_k ist $\mathfrak{g}_T$-meßbar, denn $\{\lambda \in [-\pi,\pi] \mid f(\lambda) = c_k\}$ ist Borelmenge.

Folglich muß für $k \neq l$

$$\int_{S_k} d\mathcal{G}_T \, (X'X)^{-1} \int_{S_l} d\mathcal{G}_T = 0 \quad \text{sein} \qquad (k,l = 1,\ldots,p).$$

$S_1,\ldots,S_p$ sind somit entweder selbst die Elemente des Spektrums bzgl. $\mathcal{G}_T(\lambda)$, oder aber man käme - wie oben gezeigt - durch getrennte Zerlegung von $S_1,\ldots,S_p$ zu denselben. $f(\lambda)$ ist also konstant auf den Elementen des Spektrums bzgl. $\mathcal{G}_T(\lambda)$. - Sind umgekehrt $S_1,\ldots,S_m$ die Elemente des Spektrums bzgl. $\mathcal{G}_T(\lambda)$ und ist $f(\lambda) = c_k$ für $\lambda \in S_k$ $(k = 1,\ldots,m)$, dann folgt:

$$\int_{-\pi}^{\pi} \frac{1}{f} d\mathcal{G}_T (X'X)^{-1} \int_{-\pi}^{\pi} f \, d\mathcal{G}_T = \sum_{k=1}^{m} \frac{1}{c_k} \int_{S_k} d\mathcal{G}_T \, (X'X)^{-1} \sum_{l=1}^{m} c_l \int_{S_l} d\mathcal{G}_T$$

$$= \sum_{k,l=1}^{m} \frac{c_l}{c_k} \int_{S_k} d\mathcal{G}_T (X'X)^{-1} \int_{S_l} d\mathcal{G}_T = \sum_{k=1}^{m} \int_{S_k} d\mathcal{G}_T (X'X)^{-1} \int_{S_k} d\mathcal{G}_T$$

$$= \sum_{k=1}^{m} \int_{S_k} d\mathcal{G}_T = \int_{S} d\mathcal{G}_T = X'X.$$

In (1.75) gilt also das Gleichheitszeichen. -

Korollar 1.2.: Unter der Voraussetzung von Theorem 1.5. gilt in der Ungleichung (1.75) das Gleichheitszeichen genau dann, wenn $f(\lambda)$ auf dem Spektrum S bzgl. $\mathcal{G}_T(\lambda)$ nur $p \leq n$ verschiedene Werte $c_1,\ldots,c_p$ annimmt und - bei $S_k := \{\lambda \in S \mid f(\lambda) = c_k\}$ $(k=1,\ldots,p)$ - $\sum_{k=1}^{p} r \left(\int_{S_k} d\mathcal{G}_T\right) = n$ ist.

Beweis: Ist die im Korrolar angegebene Bedingung erfüllt, so setzt man $A_k := (X'X)^{-1/2} \int_{S_k} d\mathcal{G}_T (X'X)^{-1/2}$ $(k = 1,\ldots,p)$. Genau wie im ersten Teil des Beweises von Korollar 1.1. folgt dann für $k \neq l$ $\int_{S_k} d\mathcal{G}_T (X'X)^{-1} \int_{S_l} d\mathcal{G}_T = 0$ $(k,l = 1,\ldots,p)$.

Jedes S_k ist somit die Vereinigung von Elementen des Spektrums, d.h. $f(\lambda)$ ist konstant auf jedem Element des Spektrums. - Gilt umgekehrt in (1.75) das Gleichheitszeichen, so ist $f(\lambda)$ konstant auf den Elementen des Spektrums. $f(\lambda)$ nimmt also auf S nur $p \leq n$ verschiedene Werte an. Sind $S_1,\ldots,S_p$ die Mengen aus S, in denen $f(\lambda)$ die p Werte annimmt, dann ist jedes S_k die Vereinigung von Elementen des Spektrums. Deshalb gilt für $k \neq l$ $\int_{S_k} d\mathcal{G}_T (X'X')^{-1} \int_{S_l} d\mathcal{G}_T = 0\,(k,l=1,\ldots,p)$ und aus (1.78) folgt

$$\sum_{k=1}^{p} r\left(\int_{S_k} d\mathcal{G}_T\right) = n. \; -$$

Theorem 1.6. und Korollar 1.2. entsprechen asymptotisch gültigen Theoremen (→ Theorem 2.4 bzw. Korollar 2.1.) von Grenander und Rosenblatt bzw. Anderson, die auch bereits in Verbindung mit Theorem 1.4. bzw. Korollar 1.1. als Entsprechungen genannt worden sind.

Einige Bemerkungen zu den Theoremen 1.5. und 1.6. erscheinen noch angebracht.

Zum einen besteht in Zeitreihenregressionen gewöhnlich gar kein beachtenswerter Unterschied zwischen $(X'\Sigma^{-1}X)^{-1}$ und $2\pi\left(\int_{-\pi}^{\pi} \frac{1}{f(\lambda)}\, d\mathcal{G}_T(\lambda)\right)^{-1}$, jedenfalls dann nicht, wenn die Störvariable einem stationären autoregressiven Prozeß [1] niederer Ordnung p (i.e. $p \ll T$) folgt (vgl. S.18) oder Ihre Spektraldichte sich durch die Spektraldichte eines derartigen Prozesses gut approximieren läßt (vgl. dazu 1.2.1.2.2.).

---

1) der ja, weil hier nur seine Autokovarianzen berücksichtigt werden, als invertierbar angesehen werden kann (s. S.18 ff).

Des weiteren sei darauf hingewiesen, daß zwar nicht in $\mathcal{J}_T(\lambda)$ - das ja durch die T Matrizen $N^T(\tau) = \sum_{t=1}^{T-\tau} x_{t+\tau}\, x_t'$, $\tau=0,\ldots,T-1$, vollständig bestimmt ist -, wohl aber in $f(\lambda)$ Information eingeht, die nicht der durch die Zeitpunkte $t=1,\ldots,T$ markierte Ausschnitt der Länge T des u_t-Prozesses, sondern erst der gesamte u_t-Prozeß liefert. Denn die Spektralverteilung $F(\lambda)$ des u_t-Prozesses und also (bei stetigem $f(\lambda)$) $f(\lambda) = dF(\lambda)/d\lambda$ ist erst durch die Gesamtheit der Autokovarianzen $\sigma(0)$, $\sigma(1)$, $\sigma(2)$,... vollständig bestimmt und nicht bereits durch die T Autokovarianzen $\sigma(0)$, $\sigma(1)$,...,$\sigma(T-1)$, die in der Stichprobe vom Umfang T auftreten. Letztere bestimmen allerdings Σ und damit die Varianz der Kleinstquadrat- wie der Aitken-Schätzung vollständig. Die Darstellung $X'\Sigma X = 2\pi \int_{-\pi}^{\pi} f(\lambda)\, d\mathcal{J}_T(\lambda)$ ist tatsächlich ja nur eine andere Schreibweise von $\sum_{\tau=-(T-1)}^{T-1} \sigma(\tau) N^T(\tau)$ und benutzt somit in $f(\lambda)$ enthaltene Information nur, soweit sie sich auf $\sigma(0)$, $\sigma(1)$,...,$\sigma(T-1)$ bezieht. In $\int_{-\pi}^{\pi} \frac{1}{f(\lambda)}\, d\mathcal{J}_T(\lambda)$ ist dies allerdings anders; $2\pi\left(\int_{-\pi}^{\pi} \frac{1}{f(\lambda)}\, d\mathcal{J}_T(\lambda)\right)^{-1}$ ist in der Regel eine Untergrenze der Varianz $V(\beta|X)$ der Kleinstquadrat-Schätzung, die die in $f(\lambda)$ enthaltene Information über alle Autokovarianzen des u_t-Prozesses auch benutzt.

Die Ungleichung (1.75) gilt natürlich nicht nur für die Spektraldichte $f(\lambda)$ des u_t-Prozesses, sondern überhaupt für jede in $-\pi\leq\lambda\leq\pi$ reelle, positive und stetige Funktion $g(\lambda)$, die $\int_{-\pi}^{\pi} e^{i\lambda\tau} g(\lambda)d\lambda=\sigma(\tau)$ für $\tau=0,\ldots,T-1$ erfüllt. Ist $\mathcal{K}_T$ die Klasse dieser Funktionen[1)], dann gilt also

$$(1.75') \quad V(\hat{\beta}|X) \geq 2\pi \sup_{g\in\mathcal{K}_T} \left(\int_{-\pi}^{\pi} \frac{1}{g(\lambda)}\, d\mathcal{J}_T(\lambda)\right)^{-1}.$$

---

1) Jede solche Funktion $g(\lambda)$ kann auch als Spektraldichte eines stationären Prozesses angesehenwerden(nötigenfalls ersetze man $g(\lambda)$ durch $\tilde{g}(\lambda) := (g(\lambda) + g(-\lambda))/2$, wenn g nicht gerade ist). $\mathcal{K}_T$ ist also die Klasse aller positiven, stetigen Spektraldichten derjenigen stationären Prozesse, deren erste T Autokovarianzen mit denen des u_t-Prozesses übereinstimmen.

Schließlich sei noch darauf eingegangen, wie weit sich für $f(\lambda)$ oder, um die schärfere Aussage (1.75') zu berücksichtigen, für eine der Funktionen $g(\lambda) \in \mathfrak{K}_T$ die Bedingung des Theorems 1.6. erfüllen läßt. Da $\mathfrak{g}_T(\lambda) = 1/2\pi \sum_{t=1}^{T} e^{-i\lambda t} x_t \; (\sum_{t=1}^{T} e^{-i\lambda t} x_t)^*$ gilt (s. Beweis von Lemma 4.5.), ist $\mathfrak{g}_T(\lambda)$ stetig. Ist an der Stelle ν $\mathfrak{g}_T(\lambda) \neq 0$, so auch in einer vollen Umgebung von ν; ν ist also Wachstumsstelle von $\mathfrak{g}_T(\lambda)$. Andererseits verlangt $\mathfrak{g}_T(\lambda) = 0$, daß $\sum_{t=1}^{T} e^{-i\lambda t} x_t = 0$ ist; diese Punkte bilden nur eine Menge vom Lebesgue-Maß Null.[1] Das Spektrum S bzgl. $\mathfrak{g}_T(\lambda)$ besteht sonach aus dem ganzen Intervall $[-\pi,\pi]$. $f(\lambda)$ bzw. $g(\lambda)$ darf also auf dem Intervall $[-\pi,\pi]$ höchstens n verschiedene Werte annehmen und muß als stetige Funktion somit überhaupt konstant sein.

Die Bedingungen von Theorem 1.6. bzw. Korollar 1.2. erweisen sich daher lediglich als äquivalente Formulierungen der Bedingung $\Sigma = \sigma(0)I$.

---

1) Für die Werte $\lambda \in [-\pi,\pi]$ mit $\sum_{t=1}^{T} e^{-i\lambda t} x_t = 0$ muß $z(\lambda) = e^{-i\lambda}$ gemeinsame Nullstelle der Polynome $\sum_{t=1}^{T} x_{jt} z^t \quad (j = 1,\ldots,n)$ sein; es gibt also tatsächlich höchstens T derartige Werte λ.

1.2.3. Schätzung nach prozeßkonformer Transformation des Modells

Folgt die Störvariable dem (stationären) autoregressiven Prozeß

$$\sum_{r=0}^{p} \delta_r u_{t-r} = \varepsilon_t \qquad (\delta_0 = 1,\ \delta_p \neq 0) \tag{1.17}$$

mit unkorrelierten ε_t (bei $0<\sigma_\varepsilon^2<\infty$), dann führt die Transformation

$$\sum_{r=0}^{p} \delta_r \mathcal{L}^r \qquad \text{(mit dem Lag-Operator } \mathcal{L}\text{: } \mathcal{L}z_t := z_{t-1})$$

(bei $T>p$) das Modell $y_t = \beta' x_t + u_t \quad (t=1,\ldots,T)$

in das klassische Modell

$$y_{\Delta t} := \sum_{r=0}^{p} \delta_r y_{t+p-r} = \beta' \sum_{r=0}^{p} \delta_r x_{t+p-r} + \sum_{r=0}^{p} \delta_r u_{t+p-r}$$

$$:= \beta' x_{\Delta t} + u_{\Delta t} \qquad (u_{\Delta t}=\varepsilon_{t+p};\ t=1,\ldots,T-p)$$

über. Die zeilenreguläre Matrix, die diese Transformation bewerkstelligt, ist die (T-p)xT-Matrix

$$Q_\Delta = \begin{pmatrix} \delta_p \cdots \delta_o & & \\ & \delta_p \cdots \delta_o & 0 \\ 0 & & \delta_p \cdots \delta_o \end{pmatrix} . \tag{1.20}$$

Es bezeichne X_Δ die (T-p)xn-Matrix mit den Zeilen $x_{\Delta t}(t=1,\ldots,T-p)$ und es werde angenommen, daß die Transformation Q_Δ zu keinem Rangabfall in X_Δ geführt hat: d.h. $r(X_\Delta) = r(Q_\Delta X) = n$.
Die beste lineare unverzerrte Schätzung von β im neuen Modell ist die Kleinstquadratsschätzung[1)]

$$\hat{\beta}_\Delta = (X_\Delta' X_\Delta)^{-1} \sum_{t=1}^{T-p} x_{\Delta t}\, y_{\Delta t} = \beta + (X_\Delta' X_\Delta)^{-1} \sum_{t=1}^{T-p} x_{\Delta t}\, y_{\Delta t} . \tag{1.81}$$

---

1) Worauf es hier ankommt, ist: bei gegebenem X, nicht bei gegebenem X_Δ. Denn das Relevanzmodell ist in diesem Zusammenhang $y = X\beta + u$.

Sie ist (bei gegebenem X) erwartungstreu und hat die Varianz

(1.82) $$V(\hat{\beta}_{\Delta}|X) = \sigma_{\varepsilon}^2 (X'_{\Delta} X_{\Delta})^{-1}.$$

Die Tatsache, daß man sich durch die Transformation Q_{Δ} von der Autokorrelation der Störvariablen befreit hat und die Kleinstquadratschätzung nunmehr - im transformierten Modell - wieder beste Schätzung ist, ändert aber nichts daran, daß man auf linearem Wege besser hätte schätzen können. Denn $\hat{\beta}_{\Delta}$ ist eine in y lineare Schätzung der Form C'y mit $C' = (X'_{\Delta} X_{\Delta})^{-1} X'_{\Delta} Q_{\Delta}$ und somit im Varianzmaß ungenauer als die Aitken-Schätzung (s. 1.2.1.1.). Tatsächlich ergibt sich ja auch, wenn man (1.25) mittels (4.1.5) invertiert [1]:

(1.83) $$(X'\Sigma^{-1}X)^{-1} = \sigma_{\varepsilon}^2 (X'_{\Delta} X_{\Delta})^{-1}$$
$$- \sigma_{\varepsilon}^4 (X'_{\Delta} X_{\Delta})^{-1} X'_{(1)} \Sigma_{11}^{-1/2} \{ I + \sigma_{\varepsilon}^2 \Sigma_{11}^{-1/2} X_{(1)} (X'_{\Delta} X_{\Delta})^{-1} X'_{(1)} \Sigma_{11}^{-1/2} \}^{-1} \Sigma_{11}^{-1/2} X_{(1)} (X'_{\Delta} X_{\Delta})^{-1}$$

(mit $X'_{(1)} = [x_1, \ldots, x_p]$, $\Sigma_{11} = (E(u_t u_s|X))_{t,s=1,\ldots,p}$).

Der zweite Term der rechten Seite ist negativ-semidefinit. Nach unten läßt er sich durch $-\sigma_{\varepsilon}^4 (X'_{\Delta} X_{\Delta})^{-1} X'_{(1)} \Sigma_{11}^{-1} X_{(1)} (X'_{\Delta} X_{\Delta})^{-1}$ und unter Benutzung von (1.30) somit durch

$$-2p\sigma^4 \max_{\varepsilon t=1,\ldots,p} \sigma^{tt} \cdot (X'_{\Delta} X_{\Delta})^{-1} X'_{(1)} X_{(1)} (X'_{\Delta} X_{\Delta})^{-1}$$

abschätzen (mit σ^{tt} als t-tem Diagonalelement von Σ_{11}^{-1}). Daher ist

(1.84) $$\sigma_{\varepsilon}^2 (X'_{\Delta} X_{\Delta})^{-1} \geq (X'\Sigma^{-1}X)^{-1}$$
$$\geq \sigma_{\varepsilon} (X'_{\Delta} X_{\Delta})^{-1/2} \left[I - 2p\sigma_{\varepsilon}^2 \max_{t=1,\ldots,p} \sigma^{tt} (X'_{\Delta} X_{\Delta})^{-1/2} X'_{(1)} X_{(1)} (X'_{\Delta} X_{\Delta})^{-1/2} \right] \sigma_{\varepsilon} (X'_{\Delta} X_{\Delta})^{-1/2}.$$

Die folgende heuristische Betrachtung zeigt nun, daß im Zeitreihenmodell - und zwar insbesondere bei mit polynomialem Trend behafte-

---

1) sofern der Prozeß (1.17) also invertierbar ist

ten Regressoren (zu denen auch die Scheinvariable gerechnet werden kann) - die Abweichung von $V(\hat{\beta}_\Delta | X)$ von der Varianz des Aitken-Schätzers in vielen Fällen doch nur gering ist. Setzt man nämlich

$$m_{\Delta j} := \left(\sum_{t=1}^{T-p} x_{\Delta,jt}^2 \right)^{1/2}, \quad r_{\Delta jk} := \sum_{t=1}^{T-p} x_{\Delta,jt} x_{\Delta,kt} / m_{\Delta j} m_{\Delta k} \quad (j,k=1,\ldots,n),$$

$S_\Delta := \mathrm{diag}(m_{\Delta j})_{j=1,\ldots,n}$, $R_\Delta := (r_{\Delta,jk})_{j,k=1,\ldots,n}$ und bezeichnet die Elemente von R_Δ^{-1} mit r_Δ^{jk} $(j,k=1,\ldots,n)$, dann gilt $X_\Delta' X_\Delta = S_\Delta R_\Delta S_\Delta$ und daher

$$\mathrm{Sp}\, X_{(1)} (X_\Delta' X_\Delta)^{-1} X_{(1)}' = \sum_{t=1}^{p} \sum_{j,k=1}^{n} \frac{r_\Delta^{jk}}{m_{\Delta j} m_{\Delta k}} x_{jt} x_{kt} = \sum_{j,k=1}^{n} \frac{r_\Delta^{jk}}{m_{\Delta j} m_{\Delta k}} \sum_{t=1}^{p} x_{jt} x_{kt}$$

$$\leq \left(\sum_{j=1}^{n} \sqrt{r_\Delta^{jj}} \sqrt{\frac{\sum_{t=1}^{p} x_{jt}^2}{m_{\Delta j}}} \right)^2 .$$

Ist nun (bei $p \ll T$) $\sum_{t=1}^{p} x_{jt}^2 \ll \sum_{t=1}^{T} x_{jt}^2$ und $\sum_{t=1}^{T-p} x_{\Delta,jt}^2$ etwa von derselben Größenordnung wie $\sum_{t=1}^{T} x_{jt}^2$, dann ist also, sofern die r_Δ^{jj} nicht allzu groß sind, $S_p (X_\Delta' X_\Delta)^{-1/2} X_{(1)}' X_{(1)} (X_\Delta' X_\Delta)^{-1/2} = S_p X_{(1)}' (X_\Delta' X_\Delta)^{-1} X_{(1)}$ klein. Alle Elemente von $(X_\Delta' X_\Delta)^{-1/2} X_{(1)}' X_{(1)} (X_\Delta' X_\Delta)^{-1/2}$ sind dann (betragsmäßig) klein. Berücksichtigt man noch, daß $\sigma_\varepsilon^2 = \sum_{r=0}^{p} \delta_r \sigma(r) \leq \sigma^2 \sum_{r=0}^{p} |\delta_r|$ ist ([Anderson], S.174) und $\max_{t=1,\ldots,p} \sigma^{tt} = 1/\sigma^2 \max_{t=1,\ldots,p} \omega^{tt}$ (mit den Diagonalelementen ω^{tt} von $\Omega^{-1} = (\frac{1}{\sigma^2} \Sigma)^{-1}$, so mag tatsächlich die rechte Seite von (1.84) nur wenig von $\sigma_\varepsilon^2 (X_\Delta' X_\Delta)^{-1}$ abweichen.

Läßt sich die Datenreihe x_{jt} (für $t=1,\ldots,T$) recht gut durch ein Polynom in t vom Grad m beschreiben, so ersieht man aus der Dar-

stellung

$$x_{\Delta,jt} = \sum_{r=0}^{p} \delta_r x_{j,t-p-r} = (1+\sum_{r=1}^{p} \delta_r)x_{j,t-p} - \sum_{r=1}^{p} \delta_r (x_{j,t+p} - x_{j,t+p-r}),$$

daß $\sum_{t=1}^{T-p} x^2_{\Delta,jt} \triangleq (1+\sum_{r=1}^{p} \delta_r)^2 \sum_{t=p+1}^{T} x^2_{jt}$ und also (bei $p<<T$) ebenso wie $\sum_{t=1}^{T} x^2_{jt}$ ein $O(T^{2m+1})$ ist [1]. Führt man bzgl. der x_{jt} die analogen Größen m_j und r^{jj} $(j=1,\ldots,n)$ ein, so mag zwar durch Bildung der gewogenen Differenzen p-ter Ordnung $x_{\Delta,jt}$ $(t=1,\ldots,T-p)$ die Varianz abgenommen haben, i.e. $m_{\Delta j} < m_j$ sein, auf der anderen Seite mag aber als Ausdruck reduzierter Kollinearität der transformierten Datenreihen $r^{jj}_{\Delta} < r^{jj}$ sein [2]. Die obige Argumentationsweise ist dann nicht unbegründet, wenn man davon ausgehen kann,

daß nicht allein $\sum_{t=1}^{p} x^2_{jt} \;/\; \sum_{t=1}^{T} x^2_{jt}$ $(j=1,\ldots,n)$, sondern sogar

$$\sum_{j=1}^{n} \sqrt{r^{jj}} \sqrt{\frac{\sum_{t=1}^{p} x^2_{jt}}{m_j}} \quad \text{klein ist.}$$

Zur Illustration werde wiederum der autoregressive u_t-Prozeß erster Ordnung herangezogen: $u_t = \rho u_{t-1} + \varepsilon_t$. In diesem Fall ist $\Sigma_{11} = \sigma^2$, $X'_{(1)} = x_1$, $\delta_1 = -\rho$ und $\sigma^2_\varepsilon = (1-\rho^2)\sigma^2$. Daher gilt nach (1.83)

$$(1.85) \qquad (X'\Sigma^{-1}X)^{-1} = \sigma^2_\varepsilon (X'_\Delta X_\Delta)^{-1}$$

$$- \sigma^2_\varepsilon (1-\rho^2)\{1+(1-\rho^2)x'_1(X'_\Delta X_\Delta)^{-1}x_1\}^{-1}(X'_\Delta X_\Delta)^{-1}x_1 x'_1 (X'_\Delta X_\Delta)^{-1}.$$

Den obigen Argumentationsversuch zugunsten von $\sigma^2_\varepsilon (X'_\Delta X_\Delta)^{-1} \triangleq (X'\Sigma^{-1}X)^{-1}$ wird man hier als durchaus angebracht ansehen, zumal der zweite

---

1) s.S.28f.

2) Ein häufig in Betracht kommender Fall, der dieses Argument stützt, dürfte der folgende sein: die δ_r sind negativ $(r=1,\ldots,p)$ und nehmen betragsmäßig mit wachsendem r ab und zwar derart, daß $1+\sum_{r=1}^{p} \delta_r > 0$ ist.

Term auf der rechten Seite den Faktor $1-\varrho^2$ enthält, der im vor allem interessierenden Fall eines ϱ-Wertes nahe 1 klein ist.

Man beachte aber, daß dies gerade dem Sonderfall $1 + \sum_{r=1}^{p} \delta_r \doteq 0$ entspricht, der oben außer acht gelassen wurde. Enthält die Regression nun ein Absolutglied mit der zugehörigen Scheinvariablen $x_{1t} \equiv 1$, dann ist $m_{\Delta 1} = \sqrt{T-1}\,(1-\varrho)$ und daher $\left(\sum_{j=1}^{n} \sqrt{r^{jj}}\, \frac{\sqrt{\sum_{t=1}^{p} x_{jt}^2}}{m_{\Delta j}} \right)^2 \geq \frac{1}{(T-1)(1-\varrho)^2}$. Das oben benutzte Argument versagt somit. Tatsächlich läßt sich zeigen[1], daß in diesem Fall insbesondere das Absolutglied mit geringer relativer Effizienz geschätzt wird.

Hier soll nur kurz die Auswirkung auf die verallgemeinerte Varianz besprochen werden. Aus

$$X'\Sigma^{-1}X = \frac{1}{\sigma_\varepsilon^2} X_\Delta' X_\Delta + \frac{1}{\sigma^2} x_1 x_1' \qquad ((1.25))$$

gewinnt man über (4.1.6')

$$\det X'\Sigma^{-1}X = \{1 + \frac{\sigma_\varepsilon^2}{\sigma^2} x_1' (X_\Delta' X_\Delta)^{-1} x_1\} \det (\frac{1}{\sigma_\varepsilon^2} X_\Delta' X_\Delta).$$

Also ist

1.86) $$\det \{\sigma_\varepsilon^2 (X_\Delta' X_\Delta)^{-1}\} = \{1 + (1-\varrho^2) x_1' (X_\Delta' X_\Delta)^{-1} x_1\} \det (X'\Sigma^{-1}X)^{-1}.$$

Unter Verwendung von (4.1.1.) zeigt man, daß

$$x_1' (X_\Delta' X_\Delta)^{-1} x_1 \geq x_{11}^2 \,/\, \sum_{t=1}^{T-1} x_{\Delta,1t}^2 = 1 \,/\, (T-1)(1-\varrho)^2$$

ist[2]. Daher ist der Skalar auf der rechten Seite von (1.86)

---

1) Vgl. [Kadiyala]

2) Man kann dies auch unmittelbar aus (4.4.4) ablesen.

mindestens gleich $1 + \frac{1+\varrho}{(1-\varrho)(T-1)}$. Demgemäß ungenau ist die diskutierte Schätzung .

1.2.4. Schätzung im Differenzenmodell

Häufig wird vorgeschlagen, in der zuletzt erörterten Situation die Modellparameter durch Anwendung der Kleinstquadrat-Methode im Modell mit ersten (absoluten) Differenzen zu schätzen.

Die Variablen des Differenzenmodells lauten

$$(1.87) \quad v_{\Delta t} = v_{t+1} - v_t = (v_{t+1} - \varrho v_t) + (1-\varrho)v_t \qquad (t=1,\ldots,T-1)$$

(mit $v_t = y_t$ bzw. x_t bzw. u_t). Der erste Teil der rechten Seite ist die jeweilige Variable, die man bei der im vorigen Abschnitt besprochenen Transformation des ursprünglichen Modells gewinnt. Mit einer gewissen Berechtigung darf man daher in der Tat vermuten, daß die Genauigkeit der Schätzung im Differenzenmodell der Genauigkeit der Schätzung des vorigen Abschnitts und nach den dort angestellten Betrachtungen somit der Genauigkeit der Aitken-Schätzung ziemlich nahekommt.[1)] Allerdings ist die Schätzung eines Absolutgliedes, die dort sehr ungenau ausfiel, jetzt vollkommen unmöglich geworden.

Will man die Rechtfertigung des Übergangs zum Differenzenmodell nicht auf die Argumentation des vorigen Abschnitts stützen, so

---

1) Man könnte diese Ansicht auch darauf zu gründen versuchen, daß die Autokorrelation im Differenzenmodell 'nahezu' verschwunden ist (s. (1.92)) und daher in diesem Modell die Kleinstquadrat-Schätzung 'nahezu' beste lineare unverzerrte Schätzung ist. Jedoch läßt sich, weil die zum Differenzenmodell führende Transformation singulär ist, ein lineares Schätzverfahren im ursprünglichen Modell nicht (jedenfalls nicht exakt) als ein ebensolches im Differenzenmodell auffassen.

kann man sie beispielsweise in der folgenden Betrachtung finden, die einen Zusammenhang zwischen den jeweiligen Abweichungen der Kleinstquadrat-Schätzungen im Niveau - wie im Differenzenmodell von der Aitken-Schätzung (im Niveaumodell) herstellt[1]:

Im vorliegenden Fall ist bekanntermaßen (mit $\Omega = \Sigma/\sigma^2$; s. S. 23)

$$(1-\varrho^2)X'\Omega^{-1}X = (1+\varrho^2)\sum_{t=2}^{T-1} x_t x_t' - \varrho\sum_{t=1}^{T-1} x_{t+1}x_t' - \varrho\sum_{t=1}^{T-1} x_t x_{t+1}' + x_1x_1' + x_Tx_T' .$$

Durch Umschreiben der rechten Seite entsteht daraus
bei X_Δ: $= [x_2-x_1,\ldots,x_T-x_{T-1}]$:

$$(1-\varrho^2)X'\Omega^{-1}X = (1-\varrho)^2X'X + \varrho X_\Delta'X_\Delta + \varrho(1-\varrho)(x_1x_1' + x_Tx_T').$$

Ganz ebenso ist bei y_Δ': $= (y_2-y_1,\ldots,y_T-y_{T-1})$

$$(1-\varrho^2)X'\Omega^{-1}y = (1-\varrho)^2X'y + \varrho X_\Delta' y_\Delta + \varrho(1-\varrho)(x_1y_1 + x_Ty_T).$$

Bezeichnet man nun die Kleinstquadrat-Schätzungen im Niveau - rsp. Differenzenmodell mit $\hat\beta$ rsp. β_Δ und die Aitken-Schätzung mit β^*, dann besagt die letzte Gleichung, daß

$$(1-\varrho^2)X'\Omega^{-1}X\beta^* = (1-\varrho)^2X'X\hat\beta + \varrho X_\Delta'X_\Delta\beta_\Delta + \varrho(1-\varrho)(x_1y_1 + x_Ty_T)$$

ist, woraus der vorletzten Gleichung zufolge

$$(1-\varrho)^2\, X'X\,(\beta^*-\hat\beta) + \varrho X_\Delta'X_\Delta(\beta^*-\beta_\Delta) - \varrho(1-\varrho)\{x_1(y_1-x_1'\beta^*) + x_T(y_T-x_T'\beta^*)\} = 0$$

gewonnen wird, $y_1 - x_1'\beta^*$ bzw. $y_T - x_T'\beta^*$ sind die Aitken-Residuen der Zeitpunkte 1 bzw. T. Deshalb ist

$$X_\Delta'X_\Delta(\beta_\Delta - \beta^*) \doteq -\frac{(1-\varrho)^2}{\varrho} X'X\,(\hat\beta - \beta^*) \tag{1.88}$$

und

$$E(\beta_\Delta - \beta^*)(\beta_\Delta - \beta^*)' \doteq \frac{(1-\varrho)^4}{\varrho^2}\,(X_\Delta'X_\Delta)^{-1}X'X\,E(\hat\beta-\beta^*)(\hat\beta-\beta^*)'X'X(X_\Delta'X_\Delta)^{-1}. \tag{1.89}$$

Die Abweichung der Kleinstquadrat-Schätzung im Differenzenmodell von der Aitken-Schätzungist also (im quadratischen Mittel) bei $\varrho \doteq 1$ bei weitem kleiner als die Abweichung der Kleinstquadrat-Schätzung im Niveaumodell von der Aitken-Schätzung, vorausgesetzt,

---

1) s. [Malinvaud] II, S.532.

die (empirische) Varianz der Regressoren hat durch Differenzenbildung nicht überaus stark abgenommen.[1)]

Diese Betrachtung bezieht sich zunächst nur auf ein homogenes Niveaumodell. Wie sich zeigen wird, bleibt sie aber auch für die eigentlichen Regressionskoeffizienten des inhomogenen Niveaumodells

$y_t = x_t'\beta + u_t = \alpha + z_t'\gamma + u_t$

wesentlich erhalten.

Denn in jedem Fall gilt - mit den Aitken-Residuen $u_1^* = y_1 - x_1'\beta^*$ bzw. $u_T^* = y_T - x_T'\beta^*$ und wenn man α_Δ etwa den Wert null gibt:

$$(1-\varrho)^2 X'X(\beta^*-\hat{\beta}) + \varrho X_\Delta' X_\Delta(\beta^*-\beta_\Delta) - \varrho(1-\varrho)(x_1 u_1^* + x_T u_T^*) = 0.$$

Die erste Zeile dieser Beziehung besagt, daß

$$(1-\varrho)^2\{T(\alpha^*-\hat{\alpha}) + T\,\bar{z}'\,(\gamma^*-\hat{\gamma})\} = \varrho(1-\varrho)(u_1^* + u_T^*) \quad \text{und also}$$

$$(1-\varrho)^2\{T\,\bar{z}\,(\alpha^*-\hat{\alpha}) + T\,\bar{z}\,\bar{z}'(\gamma^*-\hat{\gamma})\} = \varrho(1-\varrho)\,\bar{z}\,(u_1^* + u_T^*) \quad \text{ist;}$$

weiterhin besagen die restlichen Zeilen der Beziehung, daß

$$(1-\varrho)^2 T\,\bar{z}\,(\alpha^*-\hat{\alpha}) + (1-\varrho)^2 Z'Z(\gamma^*-\hat{\gamma}) + \varrho Z_\Delta' Z_\Delta(\gamma^*-\gamma_\Delta) = \varrho(1-\varrho)(z_1 u_1^* + z_T u_T^*).$$

ist. Insgesamt ergibt sich also

$$(1-\varrho)^2\, Z_*'Z_*(\gamma^* - \hat{\gamma}) + \varrho Z_\Delta' Z_\Delta\,(\gamma^* - \gamma_\Delta) = \varrho(1-\varrho)(z_{1*} u_1^* + z_{T*} u_T^*).$$

(1.88) und (1.89) bleiben daher bei Ersetzung von β durch γ, von X durch Z_* bzw. von X_Δ durch Z_Δ richtig.[2)]

---

1) so daß die Wirkung von $(1-\varrho)^4$ durch sehr große Elemente von $(X_\Delta' X_\Delta)^{-1} X'X$ kompensiert würde

2) Wie bereits früher, ist vorstehend Z_* die Datenmatrix der mittelwertbereinigten z-Variablen und $\bar{z}$ ihr Mittelwertvektor.

Nunmehr soll noch etwas genauer der Unterschied der Varianzen der Kleinstquadrat-Schätzungen im Niveau- und Differenzenmodell für den praxisrelevanten Fall der inhomogenen Niveauregression mit (empirisch) von starkem linearem Trend behafteten eigentlichen Regressoren angegeben werden.

Bei $z_{it} = a_i + q_i t + w_{it}$ $(i = 1,2,\ldots,n;\ t = 1,\ldots,T)$ mit Kleinstquadrat-Schätzungen a_i und q_i hat man [1]

$$(Z_*'Z_*)^{-1} = \frac{1}{T}\{ M_w^{-1} - \frac{1}{\frac{1}{s_t^2} + q'M_w^{-1}q} M_w^{-1} qq' M_w^{-1} \}$$

und (für $\tau \geq 0$)

$$N_*^T(\tau) = qq' \sum_{t=1}^{T-\tau} (t^*+\tau)t^* + q \sum_{t=1}^{T-\tau} (t^*+\tau)w_t' + \sum_{t=1}^{T-\tau} w_{t+\tau} t^* q' + \sum_{t=1}^{T-\tau} w_{t+\tau} w_t' .$$

Berücksichtigt man, daß

$$(Z_*'Z_*)^{-1} q = \frac{1}{Ts_t^2} \frac{1}{\frac{1}{s_t^2} + q'M_w^{-1}q} M_w^{-1} q = O(\frac{1}{T^3}) M_w^{-1} q$$

ist[2], so zeigt sich, daß für nicht zu kleines T

(1.90) $$(Z_*'Z_*)^{-1} N_*^T(\tau)(Z_*'Z_*)^{-1} \doteq (Z_*'Z_*)^{-1} \sum_{t=1}^{T-\tau} w_{t+\tau} w_t' (Z_*'Z_*)^{-1} \text{ gilt.}^{3)}$$

Man beachte, daß (1.90) in einem gewissen Gegensatz zu den Ergebnissen von 1.2.2.3. steht, wo gerade die Terme mit w_t's in $N_*^T(\tau)$ vernachlässigt wurden (T mag hier vielleicht größer als dort angesehen werden).

---

1) s. 1.2.2.3., insbesondere S.34. und (1.66)

2) Da annahmegemäß $s_{w_i}^2 = \sum_{t=1}^{T} w_{it}^2 / T$ klein ist, ist $q'M_w^{-1}q$ relativ groß.

3) Es ist also $\sum_{t=1}^{T-\tau} (t^*+\tau)t^*(Z_*'Z_*)^{-1}qq'(Z_*'Z_*)^{-1} = O(\frac{1}{T^3})M_w^{-1}qq'M_w^{-1}$.

Des weiteren ist $\sum_{t=1}^{T-\tau} (t^*+\tau)w_t = -\sum_{t=T-\tau+1}^{T} t w_t - \bar{t}\sum_{t=T-\tau+1}^{T} w_t - \tau\sum_{t=T-\tau+1}^{T} w_t$.

Zeigen die Trendresiduen w_t weiter empirisch das Verhalten eines reinen Zufallsprozesses, so daß also $\sum_{t=1}^{T-\tau} w_{t+\tau}\, w_t' \doteq 0$ ist für $\tau > 0$ [1], dann kann $V\hat{\gamma}$ im wesentlichen mit $\sigma^2 M_w^{-1}/T$ identifiziert werden:

$$(1.91) \qquad V\hat{\gamma} \doteq \frac{\sigma^2}{T} M_w^{-1} .$$

Dasselbe Resultat würde man im Fall einer unkorrelierten Störvariablen u_t erhalten. Die Stärke und die Art der Kollinearität der eigentlichen Regressoren führen hier also dazu, daß die Autokorrelation der Störvariablen sich nicht auswirkt.

Nun zur Varianz der Kleinstquadrat-Schätzung im Differenzenmodell. Dessen Störvariable hat die Autokovarianzfunktion

$$(1.92) \qquad \sigma_\Delta(\tau) = E(u_{t+1+\tau} - u_{t+\tau})(u_{t+1} - u_t) = 2\sigma(\tau) - \sigma(\tau-1) - \sigma(\tau+1) =$$

$$= -\sigma^2(1-\varrho)^2\, \varrho^{\tau-1} \qquad (\tau=1,2,\ldots); \quad \sigma_\Delta(0) = 2\sigma^2(1-\varrho).$$

Für $N_\Delta(\tau) = \sum_{t=1}^{T-1-\tau} x_{\Delta,t+\tau}\, x_{\Delta t}'$ hat man

$$N_\Delta(\tau) = (T-1-\tau)qq' + (w_T - w_{\tau+1})q' + q(w_{T-\tau} - w_1)' + \sum_{t=1}^{T-1-\tau} w_{\Delta,t+\tau}\, w_{\Delta t}' .$$

Nach Annahme über die w_t ist

$$X_\Delta' X_\Delta \doteq (T-1)qq' + 2\, W'W$$

und

$$N_\Delta(\tau) \doteq \begin{cases} (T-1-\tau)qq' & , \quad 2 \leq \tau \leq T-2 \\ (T-2)qq' + W'W, & \tau = 1 \end{cases} .$$

Daher ist (vgl. (1.66))

---

1) Für makroökonomische Zeitreihen mag dies häufig zutreffen. Vgl. dazu die Ergebnisse der empirischen Studie von [Schulmann].

$$(X_\Delta' X_\Delta)^{-1} \doteq \frac{1}{2T}\left(M_w^{-1} - \frac{1}{2(1+\frac{1}{2}q' M_w^{-1} q)} M_w^{-1} qq' M_w^{-1}\right)$$

und somit

$$(1.93)\quad V_{\gamma_\Delta} \doteq \sigma_\Delta(0)\,\frac{1}{2T} M_w^{-1} - 2(1-\varrho)^2\sigma^2 \cdot \frac{1}{2T} M_w^{-1}$$

$$\doteq (1-\varrho)\sigma^2 \frac{1}{T} M_w^{-1} .$$

In grober Näherung ist also $V_{\gamma_\Delta} \doteq (1-\varrho)V\hat{\gamma}$.
Gerade für diese Schätzsituation wird demnach die Differenzenbildung zu Recht empfohlen.

1.2.5. Approximative Aitken-Schätzung

In dem Schätzverfahren des letzten Abschnitts gelingt es, zumindest in der geschilderten Schätzsituation, auch ohne Kenntnis des Parameters ϱ des u_t-Prozesses eine Näherung der Aitken-Schätzung zu intendieren. Im allgemeinen wird man aber die unbekannten Parameter $\delta_1,\ldots,\delta_p,\sigma_\varepsilon^2$ des (invertiblen) autoregressiven Störvariablen-Prozesses $\sum_{r=1}^{p} \delta_r u_{t-r} = \varepsilon_t$ etwa ihrerseits erst schätzen müssen, um mit Hilfe dieser Schätzungen sodann nach (1.24) eine Schätzung $\hat{\Sigma}^{-1}$ von Σ^{-1} zu gewinnen, die in der Schätzung

$$(1.94)\quad \hat{\beta} = (X'\hat{\Sigma}^{-1}X)^{-1} X'\hat{\Sigma}^{-1}y$$

zu einer Approximation der Aitken-Schätzung dient.[1] Die Prozeßgleichung $\sum_{r=0}^{p} \delta_r u_{t-r} = \varepsilon_t$ legt die folgende Schätzung von Σ^{-1} nahe: Man führt zunächst die Kleinstquadrat-Schätzung durch, deren Residuen als Daten in der Regressionsgleichung

---

1) Es genügt natürlich, Ω^{-1} oder sonst ein Vielfaches von Σ^{-1} zu schätzen.

(1.95) $$\hat{u}_t = \sum_{r=1}^{p} (-\delta_r)\, \hat{u}_{t-r} + \eta_t \qquad (t = p+1,\ldots,T)$$

Verwendung finden. Aus (1.95) gewinnt man mittels der Kleinstquadrat-Methode Schätzungen von $(-)\delta_1,\ldots,(-)\delta_p$. Benutzt man die Residuen $\hat{u}_t$ außerdem noch zu Schätzungen

(1.96) $$\hat{\sigma}(\tau) := \frac{1}{T-\tau} \sum_{t=1}^{T-\tau} \hat{u}_{t+\tau}\, \hat{u}_t \qquad (\tau = 0,\ldots,p)$$

der Autokovarianzen des u_t-Prozesses bis zur Ordnung p, so kann man mittels

(1.97) $$\hat{\sigma}_\varepsilon^2 := \hat{\sigma}(0) + \sum_{r=1}^{p} \hat{\delta}_r\, \hat{\sigma}(r)$$

und $\hat{\Sigma}_{11}^{-1} := (\hat{\sigma}(t-s))_{t,s=1,\ldots,p}^{-1}$ Schätzungen von $\sigma_\varepsilon^2 = \sum_{r=0}^{p} \delta_r\, \sigma(r)$ bzw, Σ_{11}^{-1} gewinnen und sodann (1.24) zur Schätzung von Σ^{-1} nutzen.

Andererseits gilt $\sum_{r=0}^{p} \delta_r\, \sigma(r-s) = \sum_{r=0}^{p} \delta_r\, Eu_{t-s}u_{t-r} = Eu_{t-s}\varepsilon_t = 0$ für $s > 0$. Läßt man s die Indizes $1,\ldots,p$ durchlaufen, dann ergibt sich die Beziehung

(1.98) $$\Sigma_{11} \begin{pmatrix} \delta_1 \\ \vdots \\ \delta_p \end{pmatrix} = - \begin{pmatrix} \sigma(1) \\ \vdots \\ \sigma(p) \end{pmatrix}$$

Sie legt das folgende Vorgehen nahe: Man bestimme zunächst die Schätzungen (1.96), mit deren Hilfe $\hat{\Sigma}_{11}^{-1}$ und dann $\hat{\delta}_1,\ldots,\hat{\delta}_p$ gemäß

(1.99) $$\begin{pmatrix} \hat{\delta}_1 \\ \vdots \\ \hat{\delta}_p \end{pmatrix} = - \hat{\Sigma}_{11}^{-1} \begin{pmatrix} \hat{\sigma}(1) \\ \vdots \\ \hat{\sigma}(p) \end{pmatrix} .$$

Mit $\hat{\delta}_1,\ldots,\hat{\delta}_p$ wende man des weiteren (1.97) an.

Die Schätzung (1.94) ist nichtlinear. Auf ihre Eigenschaften in endlicher Stichprobe soll hier nicht eingegangen werden, da sie sich nur schwer absehen lassen.

1.2.6. Effiziente Schätzung

Als Abschluß der Behandlung des Schätzproblems in endlicher Stichprobe sei noch die Frage beantwortet, wie genau man denn (im mittleren quadratischen Fehler) günstigstenfalls schätzen kann. Soweit der in 4.4. definierte reguläre Schätzfall vorliegt, geben Theorem 4.4. und Korollar 4.4. die verlangte Anwort. Die dort gegebenen Bedingungen und Aussagen müssen allerdings in Bedingungen und Aussagen über die gemeinsame Dichte $f_u(u) = f(u_1,\ldots,u_T)$ der latenten Variablen umgewandelt werden, um Aussagewert zu haben. Dies soll nun geschehen.

Aus $y = X\beta + u$ folgt

$$(1.100)\quad f_y(y_1,\ldots,y_T) = f_u(y_1 - \beta' x_1,\ldots,y_T - \beta' x_T).$$

Deshalb gilt

$$(1.101)\quad \frac{\partial \log f_y}{\partial \beta} = -\sum_{t=1}^{T} \frac{\partial \log f_u}{\partial u_t} x_t = -X' \frac{\partial \log f_u}{\partial u},$$

sofern $f_u(u_1,\ldots,u_T)$ nach u_t $(t=1,\ldots,T)$ differenzierbar ist.

Sind β und θ_1 die Vektoren von (funktional unabhängigen) Parametern, die f_y innerhalb einer parametrischen Klasse von Dichtefunktionen bestimmen, dann ist offenbar eine parametrische Klasse von Dichtefunktionen $f_u(u_1,\ldots,u;\theta_1)$ gegeben, innerhalb derer f_u durch θ_1 bestimmt ist.

Der reguläre Schätzfall läßt sich dann wie folgt definieren:

Für θ_1 $(\theta_1' := (\theta_{11},\ldots,\theta_{1l}))$ aus einer offenen Menge Θ_1 des R_l sei $f_u(u_1,\ldots,u_T;\theta_1)$ positiv, nach $\theta_{11},\ldots,\theta_{1l}$ partiell differenzierbar und nach $u_1,\ldots,u_T$ stetig partiell differenzierbar. Es sei $E\left[\left(\frac{\partial \log f_u(u;\theta_1)}{\partial \theta_{1j}}\right)^2;\theta_1\right] < \infty$ $(j=1,\ldots,l)$ und die Matrix $V\left(\frac{\partial \log f_u(u;\theta_1)}{\partial \theta_1};\theta_1\right)$ regulär. Ebenso sei $E\left[\left(\frac{\partial \log f_u(u;\theta_1)}{\partial u_t}\right)^2;\theta_1\right] < \infty$

$(t=1,\ldots,T)$ und die Matrix $V\left(\frac{\partial \log f_u(u;\theta_1)}{\partial u};\theta_1\right)$ regulär. Zu θ_1 aus $\textcircled{H}_1$ gebe es eine Umgebung U_{θ_1}, so daß für $\tilde{\theta}_1 \in U_{\theta_1}$ $\left|\frac{\partial f_u(u;\tilde{\theta}_1)}{\partial \theta_{1j}}\right| \le G(u;\theta_1)$ mit einer (bzgl. des Lebesgue-Maßes) integrierbaren Funktion $G(u;\theta_1)$ ist $(j=1,\ldots,l)$. Ferner sei für die Randdichten $f_t(u_t;\theta_1)$ $\partial f_t(u_t;\theta_1)/\partial u_t$ (absolut Riemann -) integrierbar und $\lim_{|u_t|\to\infty} f_t(u_t;\theta_1) = 0$. Außerdem sei die Matrix

$$V\left(\frac{\partial \log f_u}{\partial \theta_1}\right) - E\frac{\partial \log f_u}{\partial \theta_1}\left(\frac{\partial \log f_u}{\partial u}\right)' X\left(X'V\left(\frac{\partial \log f_u}{\partial u}\right)X\right)^{-1} X'E\frac{\partial \log f_u}{\partial u}\left(\frac{\partial \log f_u}{\partial \theta_1}\right)'$$

regulär.

Schließlich seien $\hat{\beta}$ bzw. $\hat{\theta}_1$ Schätzer von β bzw. θ_1 mit $E[\hat{\beta}_i^2;\theta_1,\beta] < \infty$ $(i=1,\ldots,n)$ bzw. $E[\hat{\theta}_{1j}^2;\theta_1,\beta] < \infty$ $(j=1,\ldots,l)(\theta_1 \in \textcircled{H}_1)$.

Und für $\theta_1 \in \textcircled{H}_1$ gebe es eine Umgebung von U_{θ_1}, so daß für $\tilde{\theta}_1 \in U_{\theta_1}$

$$|\hat{\beta}_i \frac{\partial f_u(u;\tilde{\theta}_1)}{\partial \theta_{1j}}|, |\hat{\theta}_{1k} \frac{\partial f_u(u;\tilde{\theta}_1)}{\partial \theta_{1j}}|, |\hat{\beta}_i \frac{\partial f_u(u;\tilde{\theta}_1)}{\partial u_t}|, |\theta_{1k} \frac{\partial f_u(u;\tilde{\theta}_1)}{\partial u_t}| \le \tilde{G}(u;\theta_1)$$

mit einer (bzgl. des Lebesgue-Maßes) integrierbaren Funktion $\tilde{G}(u;\theta_1)$ ist $(i=1,\ldots,n;\ j,k=1,\ldots,l;\ t=1,\ldots,T)$.[1)]

---

1) Die letzten Bedingungen stellen (bei Verwendung von (1.101)) sicher, daß (4.4.2) erfüllt ist. $|\partial f_u(u;\tilde{\theta}_1)/\partial \theta_1| \le G(u;\theta_1)$ stellt (4.4.1) hinsichtlich θ_1 sicher. -
Im Beweisgang des Theorems 4.4. spielt aber nicht die Bedingung (4.4.1) selbst, sondern nur ihre Konsequenz $E[\partial \log f_y/\partial\theta;\theta] = 0$ eine Rolle. Hinsichtlich β wird dies nun (im Lichte von (1.101)) gerade durch die Bedingungen bzgl. der f_t erreicht; denn $\left|\lim_{U_1,\ldots,U_T\to\infty} \int_{-U}^{U}\ldots\int \frac{\partial f_u}{\partial u_t}\,du\right| =$

$$= \left|\lim_{U_1,\ldots,U_T\to\infty} \int_{-U_1}^{U_1}\ldots\left|\ldots\int_{-U_T}^{U_T}[f_u(u_1,..,U_t,..,u_T)-f_u(u_1,..,-U_t,..,u_T)]\,du_1..\right|^{t}..du_T\right| \le$$

$$\le \lim_{U_t\to\infty} (f_t(U_t) + f_t(-U_t)).$$

Aus Korollar 4.4. folgt nunmehr

Theorem 1.7. Im regulären Schätzfall gilt

$$(1.102)\quad E(\hat{\beta}-\beta)(\hat{\beta}-\beta)' \geq \frac{\partial E\hat{\beta}}{\partial \beta}\left(X'V\left(\frac{\partial \log f_u}{\partial u}\right)X\right)^{-1}\left(\frac{\partial E\hat{\beta}}{\partial \beta}\right)' + \left(\frac{\partial E\hat{\beta}}{\partial \theta_1} - \frac{\partial E\hat{\beta}}{\partial \beta} V_{22}^{-1} V_{12}'\right) V^{11} \left(\frac{\partial E\hat{\beta}}{\partial \theta_1} - \frac{\partial E\hat{\beta}}{\partial \beta} V_{22}^{-1} V_{12}'\right)'$$

mit $V_{22} = X'V\left(\frac{\partial \log f_u}{\partial u}\right)X$, $V_{12}' = E\,\frac{\partial \log f_y}{\partial \beta}\left(\frac{\partial \log f_y}{\partial \theta_1}\right)' = -X'E\,\frac{\partial \log f_u}{\partial u}\left(\frac{\partial \log f_u}{\partial \theta_1}\right)'$

und $V^{11} = \left\{V\left(\frac{\partial \log f_u}{\partial \theta_1}\right) - E\,\frac{\partial \log f_u}{\partial \theta_1}\left(\frac{\partial \log f_u}{\partial u}\right)' X\left(X'V\left(\frac{\partial \log f_u}{\partial u}\right)X\right)^{-1} X'E\,\frac{\partial \log f_u}{\partial u}\left(\frac{\partial \log f_u}{\partial \theta_1}\right)'\right\}^{-1}$.

Im Falle normalverteilter u folgt aus

$\log f_u = -\frac{T}{2}\log 2\pi - \frac{1}{2}\log(\det \Sigma) - \frac{1}{2}u'\Sigma^{-1}u$

$\partial \log f_u/\partial u = -\Sigma^{-1}u$, d.h. $V\left(\frac{\partial \log f_u}{\partial u}\right) = \Sigma^{-1}$. In diesem Fall stellt also $\frac{\partial E\hat{\beta}}{\partial \beta}(X'\Sigma^{-1}X)^{-1}\left(\frac{\partial E\hat{\beta}}{\partial \beta}\right)'$ eine Untergrenze von $E(\hat{\beta}-\beta)(\hat{\beta}-\beta)'$ dar, die aber nicht angenommen werden kann.

Beispielsweise sei nämlich $u_t = \varrho u_{t-1} + \varepsilon_t$ mit unabhängig und identisch $(0,\sigma_\varepsilon^2)$- normalverteilten ε_t. Die Parameterklasse, die die Verteilung von y festlegt, sei durch β (mit $-\infty<\beta_i<+\infty$, $i=1,\ldots,n$), σ^2 (mit $0<\sigma^2<\infty$) und ϱ (mit $0<\varrho<1$) angegeben. Offensichtlich sind dann hinsichtlich f_y die obigen Bedingungen des regulären Schätzfalls erfüllt. (1.23), (1.21) und (1.20) entnimmt man, daß $\det\Sigma = \sigma^2(\sigma_\varepsilon^2)^{T-1} = (\sigma^2)^T(1-\varrho^2)^{T-1}$ ist und

$$\log f_u = -\frac{T}{2}\log 2\pi - \frac{T}{2}\log\sigma^2 - \frac{T-1}{2}\log(1-\varrho^2) - \frac{1}{2\sigma^2(1-\varrho^2)}\{\varrho^2\sum_{t=2}^{T-1}u_t^2 - 2\varrho\sum_{t=1}^{T-1}u_{t+1}u_t + \sum_{t=1}^{T}u_t^2\}.$$

Es ist $\partial \log f_u/\partial\sigma^2 = \frac{1}{2\sigma^2}\left(\frac{1}{\sigma^2}u'\Omega^{-1}u - T\right)$ und folglich

$E\,\frac{\partial \log f_u}{\partial u}\,\frac{\partial \log f_u}{\partial \sigma^2} = 0$, weil hierin nur erste und dritte Momente der u_t auftreten. Weiter ist $\partial \log f_u/\partial\varrho = \frac{(T-1)\varrho}{1-\varrho^2} - \frac{1}{2\sigma^2}u'\left(\frac{\partial\Omega^{-1}}{\partial\varrho}\right)u$

und mit derselben Begründung auch $E \frac{\partial \log f_u}{\partial u} \frac{\partial \log f_u}{\partial \varphi} = 0$.
Es ist also $V'_{12} = 0$.

Als Untergrenze von $E(\hat{\beta}-\beta)(\hat{\beta}-\beta)'$ ergibt sich also bei regulärer Schätzung

$$\frac{\partial E\hat{\beta}}{\partial \beta} (X'\Sigma^{-1}X)^{-1} (\frac{\partial E\hat{\beta}}{\partial \beta})' + \frac{\partial E\hat{\beta}}{\partial \theta_1} V^{-1} (\frac{\partial \log f_u}{\partial \theta_1}) (\frac{\partial E\hat{\beta}}{\partial \theta_1})'$$

mit $\theta'_1 = (\varphi, \sigma^2)$.[1)]

Im übrigen ist Theorem (1.7) zu entnehmen, daß Kollinearität der Regressoren nicht allein, wie in 1.2.1.3. gezeigt, in der linearen Schätzklasse, sondern auch in der Klasse aller regulären Schätzer zu prinzipieller Ungenauigkeit in der Schätzung von β führt: Es ist nämlich

$$E(\hat{\beta}-\beta)(\hat{\beta}-\beta)' \geq (X'V(\frac{\partial \log f_u}{\partial u}) X)^{-1} \geq \frac{1}{\lambda_1} (X'X)^{-1} ,$$

wenn λ_1 größten Eigenwert von $V(\frac{\partial \log f_u}{\partial u})$ bezeichnet.

---

1) Dabei ist noch zu bemerken, daß $E \frac{\partial \log f_u}{\partial \sigma^2} \frac{\partial \log f_u}{\partial \varphi} \neq 0$ ist. Ferner ist für ein vernünftiges Schätzverfahren mindestens $\partial E\hat{\beta}/\partial \varphi \neq 0$.

2. Schätzung in großer Stichprobe

2.1. Voraussetzungen

Das Modell (0.1) möge nunmehr dahingehend modifiziert werden, daß zum einen $y_t = x_t'\beta + u_t$ (mit von x_t unabhängigem β), für jeden Zeitpunkt $t=1,2,\ldots$ gilt und zum anderen für (fast) jede Realisation $(x_1,x_2,\ldots)$ des x_t-Prozesses $E(u_t|x_1,x_2,\ldots) = 0$ und $E(u_t u_s|x_1,x_2,\ldots) = \sigma(t-s)$ ist $(t,s=1,2,\ldots)$; dabei sei $0<\sigma^2 = \sigma(0) < \infty$, $\sigma(\tau)$ unabhängig von $(x_1,x_2,\ldots)$ $(\tau=0,1,2,\ldots)$ und für jedes $T=1,2,\ldots$ sei $\Sigma = (\sigma(t-s))_{t,s=1,\ldots,T}$ regulär.[1)]

A.1. (i) Die Folge $\ldots,\sigma(-1),\ \sigma(0),\ \sigma(1),\ldots$ habe eine stetige Spektraldichte $f(\lambda)$.

(ii) $f(\lambda)$ sei positiv. $(-\pi\leq\lambda\leq\pi)$

Die Annahme A.1.(i) impliziert (s. [Doob], Ch.X, Sect.8), daß der u_t-Prozeß sich darstellen läßt gemäß

$$(2.1)\qquad u_t = \sum_{s=-\infty}^{\infty} \gamma_s\, \varepsilon_{t-s}$$

mit $\sum_{s=-\infty}^{\infty} \gamma_s^2 < \infty$ (γ_s reell), wobei die ε_r $E(\varepsilon_r|x_1,x_2,\ldots) = 0$, $E(\varepsilon_r\varepsilon_s|x_1,x_2,\ldots) = \delta_{rs}\,\sigma_\varepsilon^2$ mit $0<\sigma_\varepsilon^2<\infty$ erfüllen (σ_ε^2 und die γ_s hängen von $(x_1,x_2,\ldots)$ nicht ab). Es gilt dann gerade

$$f(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi}\Big|\sum_{s=-\infty}^{\infty} \gamma_s\, e^{i\lambda s}\Big|^2 \qquad (-\pi\leq\lambda\leq\pi).$$

Die Annahme A.1. ist beispielsweise für stationäre autoregressive u_t-Prozesse erfüllt (seien sie nun invertibel oder nicht).

---

1) Die Voraussetzungen des Modells (0.1) sind in diesen Annahmen enthalten (s.Fn.2, S.11). - Beispielsweise möge der Prozeß $\{u_t\}$ schwach stationär sein und unabhängig vom x_t-Prozeß (d.h. irgend endlich viele u_t und x_s sind stochastisch unabhängig).

A.1. (i) wird im Hinblick auf asymptotische Normalverteilung der Schätzung von β verschärft zu

A.1.' $u_t = \sum_{s=-\infty}^{\infty} \gamma_s \varepsilon_{t-s}$ mit unabhängig und identisch verteilten ε_s,

$$E\varepsilon_t = 0, \quad 0 < E\,\varepsilon_t^2 < \infty \quad \text{und} \quad \sum_{s=-\infty}^{\infty} |\gamma_s| < \infty$$

(bei gegebener Realisation des x_t-Prozesses).

Unter der Annahme A.1.' ist aufgrund der gleichmäßigen Konvergenz von $|\sum_{s=-\infty}^{\infty} \gamma_s e^{i\lambda s}|$ $f(\lambda)$ von selbst stetig $(-\pi \leq \lambda \leq \pi)$.

A.2. (i) Es gebe Funktionen $k_1(T) \uparrow \infty, \ldots, k_n(T) \uparrow \infty$ derart, daß für jedes $\tau = 0,1,2,\ldots$ für (fast) jede Realisation des x_t-Prozesses

$m_{ij}^{T}(\tau) := \sum_{t=1}^{T-\tau} x_{it+\tau} x_{jt} / k_i(T) k_j(T)$ für $T \to \infty$ gegen eine Zahl $m_{ij}(\tau)$ konvergiert $(i,j = 1,\ldots,n)$. Es sei $0 < m_{ii}(0) < \infty$ $(i = 1,\ldots,n)$ und die Matrix $M(0) := (m_{ij}(0))_{i,j=1,\ldots,n}$ sei regulär.

(ii) Es gelte $\lim_{T\to\infty} k_i(T-1)/k_i(T) = 1 \qquad (i = 1,\ldots,n)$.

Bezeichnet man mit K_T die Diagonalmatrix

(2.2) $K_T := \operatorname{diag}(k_1(T),\ldots,k_n(T))$,

so folgt aus A.2.

(2.3) $\lim_{T\to\infty} \max_{t \leq T} |K_T^{-1} x_t| = 0$

Denn[1] es gilt für $i = 1,\ldots,n$

---

1) s. [Malinvaud] II, S.226f.

$$\frac{x_{i\,T}^2}{k_i^2(T)} = \frac{\sum_{t=1}^{T} x_{i\,t}^2}{k_i^2(T)} - \frac{\sum_{t=1}^{T-1} x_{i\,t}^2}{k_i^2(T-1)} + \frac{\sum_{t=1}^{T-1} x_{i\,t}^2}{k_i^2(T-1)} \left(1 - \frac{k_i^2(T-1)}{k_i^2(T)}\right) \longrightarrow 0 .$$

Gälte nun $\max_{t\le T} x_{i\,t}^2 / k_i^2(T) \not\to 0$, so gäbe es ein $\varepsilon_o > 0$ und eine Indexfolge $T_s \uparrow \infty$ derart, daß $x_{i\,\tau_s}^2 / k_i^2(T_s) = \max_{t\le T_s} x_{i\,t}^2 / k_i^2(T_s) \ge \varepsilon_o$ wäre. Dann gälte $\tau_s \to \infty$ für $s\to\infty$ und $x_{i\,\tau_s}^2 / k_i^2(\tau_s) \ge x_{i\,\tau_s}^2 / k_i^2(T_s) \ge \varepsilon_o$, während doch $x_{i\,\tau_s}^2 / k_i^2(\tau_s)$ für $s\to\infty$ gegen null streben müßte.

Die Existenz einer Folge $k_i(T)$ mit $k_i(T)\uparrow\infty$, $k_i(T-1)/k_i(T)\to 1$ und $\sum_{t=1}^{T} x_{i\,t}^2 / k_i^2(T) \to m_{ii}(0)$ mit $0<m_{ii}(0)<\infty$ ist gleichbedeutend mit der Bedingung

$$(2.4) \qquad \max_{s\le T} x_{i\,s}^2 \Big/ \sum_{t=1}^{T} x_{i\,t}^2 \to 0.$$

Denn $\max_{s\le T} x_{i\,s}^2 \Big/ \sum_{t=1}^{T} x_{i\,t}^2 = (\max_{s\le T} x_{i\,s}^2 / k_i^2(T)) / m_{ii}^T(0) \to 0$; umgekehrt folgt aus (2.4) (wenn nur ein $x_{it} \neq 0$ ist) $\sum_{t=1}^{T} x_{i\,t}^2 \uparrow \infty$ und man kann $k_i(T) = \sum_{t=1}^{T} x_{i\,t}^2$ wählen.

Die Annahme A.2. ist recht allgemein, zumindestim Teil (i), der eine gewisse (starke) Ergodizitätsforderung an den x_t-Prozeß hinsichtlich dessen zweiter (Reihen-)Momente stellt.

Im folgenden wird die Matrix $(m_{ij}^T(\tau))_{i,j=1,\ldots,n}$ mit $M^T(\tau)$ bezeichnet, entsprechend die Matrix $(m_{ij}(\tau))_{i,j=1,\ldots,n}$ mit $M(\tau)$ und für $\tau<0$ $m_{ij}^T(\tau) = m_{ji}^T(-\tau)$, $m_{ij}(\tau) = m_{ji}(-\tau)$ gesetzt $(-(T-1) \le \tau < 0$; $i,j=1,\ldots,n)$. Für $|\tau| \ge T$ werde $M^T(\tau) = 0$ gesetzt. Es gilt dann

$M^T(-\tau)=M^T(\tau))'$, $M(-\tau) = (M(\tau))'$ $(\tau=0,\pm1,\ldots)$ und offenbar ist

$$(2.5)\qquad M^T(\tau) = K_T^{-1}\, N^T(\tau)\, K_T^{-1} \qquad (\tau=0,\pm1,\ldots).$$

2.2. Asymptotische Varianz von Kleinstquadrat- und Aitken-Schätzung

Setzt man mit der Spektralverteilungsfunktion $\mathcal{G}_T(\lambda)$ der Folge $\{N^T(\tau)\}_{\tau=0,\pm1,\ldots}$

$$(2.6)\qquad \mathcal{F}_T(\lambda) := K_T^{-1}\mathcal{G}_T(\lambda)K_T^{-1} \qquad (-\infty<\lambda<\infty),$$

dann ist $\mathcal{F}_T(\lambda)$ die Spektralverteilungsfunktion der Folge $\{M^T(\tau)\}_{\tau=0,\pm1,\ldots}$. Für die Varianz der Kleinstquadrat-Schätzung $\hat{\beta}$ gilt somit nach (1.47) und (1.49')

$$(2.7)\qquad K_T V(\hat{\beta}) K_T = 2\pi\, M^{T-1}(0) \int_{-\pi}^{\pi} f(\lambda)\, d\mathcal{F}_T(\lambda)\, M^{T-1}(0)$$

und man bekommt das folgende

Theorem 2.1. (Grenander und Rosenblatt)[1] Unter den Voraussetzungen A.1.(i) und A.2.(i) gilt für die Varianz der Kleinstquadrat-Schätzung (bei gegebener Realisation des x_t-Prozesses)

$$(2.8)\qquad \lim_{T\to\infty} K_T V(\hat{\beta}) K_T = 2\pi\, M^{-1}(0) \int_{-\pi}^{\pi} f(\lambda)\, d\mathcal{F}(\lambda)\, M^{-1}(0)$$

Darin ist $\mathcal{F}(\lambda)$ die Spektralverteilungsfunktion der Folge $\{M(\tau)\}_{\tau=0,\pm1,\ldots}$.

Beweis: Die Folge $\{M(\tau)\}_{\tau=0,\pm1,\ldots}$ hat eine Spektralverteilungsfunktion $\mathcal{F}$ und es gilt $\mathcal{F}_T(\lambda) \longrightarrow \mathcal{F}$ an jeder Stetigkeitsstelle λ von $\mathcal{F}$ (Theorem 4.5.1.).

---

1) [Grenander und Rosenblatt], S.240 haben zur Herleitung allerdings noch als weitere Voraussetzung A.2.(ii) benutzt.

Wegen der Stetigkeit von $f(\lambda)$ ergibt sich die Behauptung unmittelbar aus (2.7). -

Theorem 2.2. (Grenander und Rosenblatt)[1] Unter den Voraussetzungen A.1. und A.2. gilt für die Varianz der Aitken-Schätzung β^* (bei gegebener Realisation des x_t-Prozesses)

$$(2.9) \quad \lim_{T\to\infty} K_T V(\beta^*) K_T = 2\pi \left(\int_{-\pi}^{\pi} \frac{1}{f(\lambda)} d\mathcal{F}(\lambda) \right)^{-1}.$$

Beweis: Ist der u_t-Prozeß autoregressiv und invertibel, so folgt mittels (2.3) aus Theorem 1.1. unmittelbar

$$K_T^{-1} X' \Sigma^{-1} X K_T^{-1} \longrightarrow \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{f(\lambda)} d\mathcal{F}.$$

Andernfalls gibt es, da $M^T(0)$ (elementweise) beschränkt ist, zu jedem $\varepsilon>0$ zwei invertible autoregressive Prozesse, deren Spektraldichten f_1, f_2 $\left|\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{f_1} d\mathcal{F}_T - \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{f_2} d\mathcal{F}_T\right| < \varepsilon$ erfüllen und für deren Varianzmatrizen Σ_1, Σ_2

$$(2.10) \quad K_T^{-1} X' \Sigma_2^{-1} X K_T^{-1} \leq K_T^{-1} X' \Sigma^{-1} X K_T^{-1} \leq K_T^{-1} X' \Sigma_1^{-1} X K_T^{-1}$$

gilt ((1.32), (1.36) und (1.34)). Nach dem bereits Bewiesenen konvergieren die rechte bzw. linke Seite von (2.10) gegen $\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{f_i} d\mathcal{F}$ (i=1,2), und die Limesmatrizen unterscheiden sich betragsmäßig höchstens um ε. Des weiteren wendet man (1.35) an. Da M(0) regulär ist, ist auch $\int_{-\pi}^{\pi}\frac{1}{f} d\mathcal{F}$ regulär. -

Die Beweise von Theorem 2.1. sowohl als auch von Theorem 2.2. sind hier auf anderem Wege geführt als in der zitierten Quelle.

Für den invertiblen autoregressiven u_t-Prozeß $\sum_{r=0}^{p} \delta_r u_{t-r} = \varepsilon_t$ entnimmt man (1.28):

$$(2.11) \quad \lim_{T\to\infty} K_T V(\beta^*) K_T = \sigma_\varepsilon^2 \left(\sum_{r,s=0}^{p} \delta_r \delta_s M(s-r) \right)^{-1}.$$

---

1) op. cit., S.240. - Für beide Theoreme genügt die stückweise Stetigkeit von $f(\lambda)$, solange die Unstetigkeitsstellen von $f(\lambda)$ nicht mit Unstetigkeitsstellen von $\mathcal{F}(\lambda)$ zusammenfallen.

Auch für $\lim K_T V(\hat{\beta}) K_T$ gewinnt man eine Darstellung, in der die 'Momentenmatrizen' $M(\tau)$ des x_t-Prozesses explizit auftreten, und zwar in dem Fall, daß $\sum_{\tau=-\infty}^{\infty} |\sigma(\tau)| < \infty$ ist. Dann ist nämlich

$f(\lambda) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \sigma(\tau) e^{i\lambda\tau}$ gleichmäßig beschränkt $(-\pi \le \lambda \le \pi)$ und in

$$\sum_{\tau=-T_1}^{T_2} \sigma(\tau)\, M(\tau) = \int_{-\pi}^{\pi} \sum_{\tau=-T_1}^{T_2} \sigma(\tau) e^{i\lambda\tau}\, d\mathcal{F}$$

strebt die rechte Seite aufgrund des Satzes von der majorisierten Konvergenz für $T_1, T_2 \to \infty$ gegen $2\pi \int_{-\pi}^{\pi} f(\lambda)\, d\mathcal{F}$; also muß auch die linke Seite konvergieren und es gilt:

$$\text{(2.12)} \qquad \lim_{T\to\infty} K_T V(\hat{\beta}) K_T = M^{-1}(0) \sum_{\tau=-\infty}^{\infty} \sigma(\tau)\, M(\tau)\, M^{-1}(0) \quad .$$

Für den Prozeß $u_t = \varrho u_{t-1} + \varepsilon_t$ beispielsweise ist hiernach

$$\lim_{T\to\infty} K_T V(\hat{\beta}) K_T = \sigma^2\, M^{-1}(0) \sum_{\tau=-\infty}^{\infty} \varrho^{|\tau|}\, M(\tau)\, M^{-1}(0),$$

während sich nach (2.11)

$$\lim_{T\to\infty} K_T V(\beta^*) K_T = \sigma_\varepsilon^2 ((1+\varrho^2) M(0) - \varrho M(1) - \varrho M(-1))^{-1}$$

ergibt.

Dafür läßt sich auch schreiben

$$\text{(2.13)} \qquad \lim_{T\to\infty} K_T V(\hat{\beta}) K_T = \sigma^2 \frac{1+\varrho}{1-\varrho} M^{-1}(0) - 2\sigma^2 M^{-1}(0) \sum_{\tau=1}^{\infty} \varrho^\tau \{ M(0) - \frac{M(\tau)+M(-\tau)}{2} \} M^{-1}(0)$$

bzw.

$$\text{(2.14)} \qquad \lim_{T\to\infty} K_T V(\beta^*) K_T = \sigma^2 \frac{1+\varrho}{1-\varrho} \left\{ M(0) + \frac{2\varrho}{(1-\varrho)^2} \left(M(0) - \frac{M(1)+M(-1)}{2} \right) \right\}^{-1} \quad 1) \quad .$$

---

1) Da die Folge $\{M(\tau)\}$ positiv-semidefinit ist, sind die Differenzen $M(0) - \frac{M(\tau)+M(-\tau)}{2}$ positiv-semidefinit symmetrisch $(\tau = 1,2,\ldots)$.

Der Beweis von Theorem 2.1., der sich auf Theorem 4.5.1. gründet, macht keinen Gebrauch davon, daß $k_i(T)$, also $\sum_{t=1}^{T} x_{it}^2$, (monoton) gegen unendlich geht (i=1,...,n). Solange nur mit einer Folge regulärer symmetrischer (n x n)-Matrizen K_T $K_T^{-1}N^T(\tau)K_T^{-1}$ gegen, sagen wir, $S(\tau)$ konvergiert (τ=0,±1,...) und S(0) regulär ist, kann bei stetigem $f(\lambda)$ $2\pi\, S^{-1}(0) \int_{-\pi}^{\pi} f(\lambda)\, d\mathfrak{F}(\lambda)\, S^{-1}(0)$ (mit der zur Folge $\{S(\tau)\}_{\tau=0,\pm1,\ldots}$ gehörigen Spektralverteilungsfunktion $\mathfrak{F}(\lambda)$) als asymptotische Varianz der Kleinstquadrat-Schätzung angesehen werden.[1)]

In jedem Falle gilt dann - gleichgültig, ob nun für ein i $\sum_{t=1}^{T} x_{it}^2$ beschränkt ist oder aber für jedes i=1,...,n die Summe $\sum_{t=1}^{T} x_{it}^2$ gegen unendlich geht, von keinem ihrer Summanden dominiert (wie in (2.4) ausgedrückt) oder nicht (d.h. (2.4) ist nicht erfüllt):

$$(2.15)\quad \lim_{T\to\infty} K_T V(\hat{\beta}) K_T = 2\pi\, S^{-1}(0) \int_{-\pi}^{\pi} f(\lambda)\, d\mathfrak{F}\, S^{-1}(0)$$

und bei $f(\lambda) > 0$ $\quad(-\pi \leq \lambda \leq \pi)$

$$(2.16)\quad \lim_{T\to\infty} K_T V(\hat{\beta}) K_T \geq 2\pi \left(\int_{-\pi}^{\pi} \frac{1}{f(\lambda)}\, d\mathfrak{F}(\lambda) \right)^{-1} .$$

Das Gleichheitszeichen in (2.16) trifft genau dann zu, wenn $f(\lambda)$ konstant auf den Elementen des Spektrums von $\mathfrak{F}$ ist.[2)]

In diesem Zusammenhang sei daran erinnert, daß in 1.2.2.5. die Klasse $\mathfrak{k}_T$ aller in $-\pi \leq \lambda \leq \pi$ stetigen, positiven und geraden Funktionen $g(\lambda)$ eingeführt worden war, die für $0 \leq \tau \leq T-1$ $\int_{-\pi}^{\pi} e^{i\lambda\tau} g(\lambda)\, d\lambda = \sigma(\tau)$ erfüllen. Es war gezeigt worden, daß

$$(1.75')\quad V(\hat{\beta}) \geq 2\pi \sup_{g \in \mathfrak{k}_T} \left(\int_{-\pi}^{\pi} \frac{1}{g(\lambda)}\, d\mathfrak{G}_T(\lambda) \right)^{-1}$$

---

1) wenn nur die Diagonalelemente dieser Matrix positiv sind.

2) Zum Beweis von (2.16) und dieser Bedingung für die Gültigkeit des Gleichheitszeichens sei auf den Beweis von Theorem 2.4. verwiesen.

ist. Die Mengen $\check{k}_T$ bilden eine absteigende Folge, und in diesem Sinne wird die letzte Ungleichung mit wachsendem T fortlaufend schwächer. Die gesamte Autokovarianzfunktion $\sigma(0)$, $\sigma(1)$,... bestimmt nun (das stetige) $f(\lambda)$; daher konvergiert die Folge der Mengen $\check{k}_T$ gegen die Menge mit dem einen Element $f(\lambda)$ und die (von links und rechts mit K_T multiplizierte) Ungleichung (1.75') reduziert sich für $T\to\infty$ auf (2.16).

Die Existenz einer asymptotischen Varianz der Kleinstquadrat-Schätzung läßt sich auch dann sicherstellen, wenn man gar keine Anforderungen an die Struktur des u_t-Prozesses stellt[1], dafür aber die Anforderungen an die Struktur (der gegebenen Realisation) des x_t-Prozesses erhöht.

Theorem 2.3. Gilt für die Folge symmetrischer Matrizen K_T

(2.17) $$\lim_{T\to\infty} S^T(\tau) = \lim_{T\to\infty} K_T^{-1} N^T(\tau) K_T^{-1} = S(\tau),$$
ist $S(0)$ regulär und konvergiert
$$\mathfrak{f}_T(\lambda) = \frac{1}{2\pi} \sum_{\tau=-(T-1)}^{T-1} e^{-i\lambda\tau} S^T(\tau) \quad \text{gegen } \mathfrak{f}(\lambda) \quad \text{(elementweise)}$$
majorisiert von einer F-integrierbaren Funktion $(-\pi\le\lambda\le\pi)$, dann ist (bei gegebener Realisation des x_t-Prozesses)

(2.18) $$\lim_{T\to\infty} K_T V(\hat{\beta}) K_T = 2\pi\, S^{-1}(0) \int_{-\pi}^{\pi} \mathfrak{f}(\lambda)\, dF(\lambda)\, S^{-1}(0).$$
($F(\lambda)$ ist die Spektralverteilungsfunktion des u_t-Prozesses).

Der Beweis ist unmittelbar (1.49) zu entnehmen. -

Will man hingegen die Existenz einer asymptotischen Varianz der Aitken-Schätzung sichern, dann bereitet es Schwierigkeiten, von

---

1) wie die in A.1.(i) ausgesprochene Forderung, daß der Prozeß linear ist und seine Spektraldichte stetig

den Voraussetzungen A.1. und A.2. abzugehen. Selbst unter der Annahme A.1. bereits kann ja, wie (1.25) - (1.28) zeigen, die Voraussetzung A.2. nicht gelockert werden (da nämlich (2.3) wesentlich ist). Will man andererseits über das Verhalten der x_t keine wesentlich andere Annahme machen als die, daß ihre asymptotischen Automomentenmatrizen $M(\tau)$ (nach geeigneter Normalisierung der $N^T(\tau)$) existieren, dann erfordert dies, daß man $X'\Sigma^{-1}X = \sum_{t,s=1}^{T} \sigma^{ts} x_t x_s'$ bis auf einen nach Normalisierung kleinen Fehler als Ausdruck in den $N^T(\tau)$ schreiben kann. Insoweit müssen die Inversen Σ^{-1} der Kovarianzmatrizen Σ sich ihrerseits als Kovarianzmatrizen eines (und desselben) stationären Prozesses auffassen lassen, was auf die autoregressiven Prozesse und damit noch auf überhaupt sämtliche Prozesse mit positiver stetiger Spektraldichte verweist.

2.3. Bedingungen für die asymptotische Effizienz der Kleinstquadrat-Schätzung

Theorem 2.4. (Grenander und Rosenblatt)[1]. Unter den Voraussetzungen A.1. und A.2. ist die Kleinstquadrat-Schätzung asymptotisch effizient genau dann, wenn $f(\lambda)$ konstant auf den Elementen des Spektrums von $\mathfrak{T}$ ist.

Unter asymptotischer Effizienz wird hier verstanden, daß $K_T V(\hat{\beta}) K_T$ und $K_T V(\beta^*) K_T$ für $T\to\infty$ dieselbe Limesmatrix haben; präziser, daß unter der Annahme A.1. die Gleichheit der Limesmatrizen für jede Folge K_T zutrifft, die die Annahme A.2. erfüllt. Es genügt aber, dies für eine einzige derartige Folge K_T zu zeigen. Denn sind K_T und $\tilde{K}_T$ zwei solche Folgen, dann gilt bei Bezeichnung der in A.2. auftretenden Limesmatrizen mit $M(\tau)$ bzw. $\tilde{M}(\tau)$:

$$k_i(T)/\tilde{k}_i(T) = (\sqrt{\Sigma x_i^2}/\tilde{k}_i(T))/(\sqrt{\Sigma x_i^2}/k_i(T)) \longrightarrow \sqrt{\tilde{m}_{ii}(0)}/\sqrt{m_{ii}(0)} := c_i > 0,$$

---

1) [Grenander und Rosenblatt], S. 244

d.h. $K_T \tilde{K}_T^{-1} = \tilde{K}_T^{-1} K_T \longrightarrow C := \mathrm{diag}(c_1,\ldots,c_n)$ und also etwa $\lim_{T\to\infty} \tilde{K}_T V(\hat{\beta}) \tilde{K}_T = C^{-1} \lim_{T\to\infty} K_T V(\hat{\beta}) K_T\, C^{-1}$. Ferner gilt $\tilde{M}(\tau) = C\, M(\tau)\, C$ $(\tau=0,\pm1,\ldots)$ und damit für die zugehörigen Spektralverteilungsfunktionen $\tilde{\mathcal{F}}(\lambda) = C\, \mathcal{F}(\lambda)\, C$. $\tilde{\mathcal{F}}$ und $\mathcal{F}$ haben daher auch dasselbe Spektrum und dieselben Elemente des Spektrums.

Nun zum Beweis des Theorems, der wiederum anders geführt wird als bei Grenander und Rosenblatt. Tatsächlich ist er in 1.2.2.5. bereits geliefert. Denn nach Theorem 4.5.2. gibt es ein (Intervall-)Maß μ, bzgl. dessen $\mathcal{F}(\lambda)$ eine Dichte $f(\lambda)$ hat. M(0) tritt an die Stelle von X'X, $\mathcal{F}$ an die Stelle von $\mathcal{G}_T$, f an die Stelle von $\mathcal{G}_T$ und μ an die Stelle des Lebesgue-Maßes im Beweis zu Theorem 1.6., der ansonsten wörtlich übernommen werden kann.

Korollar 2.1. (Anderson)[1]. Unter den Voraussetzungen A.1. und A.2. ist die Kleinstquadrat-Schätzung asymptotisch effizient genau dann, wenn $f(\lambda)$ auf dem Spektrum S von $\mathcal{F}$ nur $p \leq n$ verschiedenen Werte $c_1,\ldots,c_p$ annimmt und -
bei $S_k := \{\lambda \in S \mid f(\lambda) = c_k\}$ -

$$\sum_{k=1}^{p} r\left(\int_{S_k} d\mathcal{F}\right) = n \quad \text{ist.}$$

Beweis: Im Beweis von Korollar 1.2 ist lediglich X'X durch M(0) und $\mathcal{G}_T$ durch $\mathcal{F}$ zu ersetzen. -

Hinreichend und notwendig dafür, daß unter der Annahme A.2. die Kleinstquadrat-Schätzung asymptotisch effizient ist für alle u_t-Prozesse mit positiver stetiger Spektraldichte, ist, daß $\mathcal{F}(\lambda)$ höchstens $p \leq n$ verschiedene Wachstumsstellen $\lambda_1,\ldots,\lambda_p$ hat und

$$\sum_{k=1}^{p} r(\mathcal{F}(\lambda_k) - \mathcal{F}(\lambda_k - 0)) = n \quad \text{ist.}^{2)}$$

---

1) [Anderson], S.580.

2) [Anderson], S.581. - Die Aussage folgt unmittelbar aus Korollar 2.1.

2.4. Asymptotische Verteilung der Koeffizientenschätzungen

Theorem 2.5. (Anderson)[1]. Unter den Voraussetzungen A.1.' und A.2. gilt für die Kleinstquadrat-Schätzung $\hat{\beta}$:
$K_T(\hat{\beta}-\beta)$ konvergiert (bei gegebener Realisation des x_t-Prozesses) in Verteilung gegen eine $N(0, 2\pi M^{-1}(0) \int_{-\pi}^{\pi} f(\lambda) d\mathcal{F} M^{-1}(0))$-verteilte Variable.

Beweis[2]: Es genügt zu zeigen, daß $K_T^{-1}X'u$ in Verteilung gegen eine nach $N(0, 2\pi \int_{-\pi}^{\pi} f(\lambda)\, d\mathcal{F})$ verteilte Variable strebt. Nun ist für jeden konstanten Vektor $c \neq 0$ für $K = 1,2,\ldots$

$$(2.19)\quad c'K_T^{-1}X'u = c'K_T^{-1}\sum_{t=1}^{T} x_t u_t = c'K_T^{-1}\sum_{t=1}^{T} x_t \sum_{s=-K}^{K} \gamma_s \varepsilon_{t-s} + c'K_T^{-1}\sum_{t=1}^{T} x_t \sum_{|s|>K} \gamma_s \varepsilon_{t-s}$$

$$:= z_{KT} + w_{KT}\,.$$

Nach einem Theorem Anderson's ([Anderson], S.426) genügt es zu zeigen:

1° $Ew_{KT}^2 \leq m_K \xrightarrow[K\to\infty]{} 0$

2° z_{KT} konvergiert bei festem $K \geq K_0$ für $T\to\infty$ gegen eine nach $N(0,V_K)$ verteilte Variable

3° $V_K \xrightarrow[K\to\infty]{} 2\pi\, c' \int_{-\pi}^{\pi} f(\lambda)\, d\mathcal{F}\, c$. -

1°. Man hat

$$Ew_{KT}^2 = c'K_T^{-1} \sum_{t,t'=1}^{T} x_t x_{t'}' \sum_{|s|,|s'|>K} \gamma_s \gamma_{s'} E\varepsilon_{t-s}\varepsilon_{t'-s'} K_T^{-1} c$$

$$= \sigma_\varepsilon^2\, c'K_T^{-1} \sum_{t,t'=1}^{T} x_t x_{t'}' \sum_{\substack{|s|,|s'|>K \\ s=s+t'-t}} \gamma_s \gamma_{s'} K_T^{-1} c \quad .$$

---

1) Anderson's Voraussetzung über den u_t-Prozeß ist etwas allgemeiner, nämlich: $u_t = \sum_{s=-\infty}^{\infty} \gamma_s \varepsilon_{t-s}$ mit $\sum_{s=-\infty}^{\infty} |\gamma_s| < \infty$ und unabhängig verteilten ε_t, die $\sup_{t=1,2,\ldots} \int_{|\varepsilon|>c} \varepsilon^2 dF_{\varepsilon_t} \to 0$ für $c\to\infty$ erfüllen.

2) i.w. nach [Anderson], S.585 f.

Da $\sum_{r=-\infty}^{\infty} |\gamma_r| < \infty$ ist, kann man hierin die Summationsreihenfolge vertauschen:

$$Ew_{KT}^2 = \sigma_\varepsilon^2 \, c' K_T^{-1} \sum_{|s|,|s'|>K} \gamma_s \gamma_{s'} \sum_{\substack{1 \le t \le T \\ 1 \le t+s'-s \le T}} x_t \, x'_{t+s'-s} \, K_T^{-1} c$$

$$= \sigma_\varepsilon^2 \sum_{|s|,|s'|>K} \gamma_s \gamma_{s'} \, c' M^T(s'-s) c \le k \Big(\sum_{|s|>K} |\gamma_s| \Big)^2 \xrightarrow[K\to\infty]{} 0.$$

Denn $\sum_{\substack{|s|,|s'|>K \\ s \ne s'}} \gamma_s \gamma_{s'} \, c' M^T(s'-s) c \le \sum_{\substack{|s|,|s'|>K \\ s'>s}} |\gamma_s| |\gamma_{s'}| \cdot 2c' M^T(0) c$ [1] und

$c' M^T(0) c$ ist aufgrund der Konvergenz von $M^T(0)$ beschränkt.

$$2^{\circ}. \; z_{KT} = \sum_{r=K+1}^{T-K} \varepsilon_r \, c' K_T^{-1} \sum_{s=-K}^{K} \gamma_s \, x_{r+s}$$

$$+ \sum_{r=1-K}^{K} \varepsilon_r \, c' K_T^{-1} \sum_{s=1-r}^{K} \gamma_s x_{r+s} + \sum_{r=T-K+1}^{T+K} \varepsilon_r \, c' K_T^{-1} \sum_{s=-K}^{T-r} \gamma_s \, x_{r+s} \qquad (T > 2K).$$

Die Varianz jedes der letzten beiden Ausdrücke ist durch

$4K^2 \sigma_\varepsilon^2 \max_{|s| \le K} |\gamma_s|^2 \; |c|^2 \max_{t \le T} |K_T^{-1} x_t|$ beschränkt; aus (2.3) folgt, daß beide Ausdrücke bei festem K für $T \to \infty$ in Wahrscheinlichkeit gegen null gehen. Weiterhin ist $\sum_{r=K+1}^{T-K} \varepsilon_r \, c' K_T^{-1} \sum_{s=-K}^{K} \gamma_s \, x_{r+s}$ von der Gestalt $\sum_{r=K+1}^{T-K} \varepsilon_r \, a_{rT}$, wobei nach (2.3) $\max_{K+1 \le r \le T-K} |a_{rT}| \xrightarrow[T\to\infty]{} 0$ und

$$\sum_{r=K+1}^{T-K} a_{rT}^2 = \sum_{r=K+1}^{T-K} c' K_T^{-1} \sum_{s,s'=-K}^{K} \gamma_s \gamma_{s'} \, x_{r+s} \, x'_{r+s'} \, K_T^{-1} c$$

---

1) s. (1.51)

$$= \sum_{s,s'=-K}^{K} \gamma_s \gamma_{s'}\, c' K_T^{-1} \sum_{r=K+1}^{T-K} x_{r+s}\, x'_{r+s'}\, K_T^{-1} c$$

$$\xrightarrow[T\to\infty]{} \sum_{s,s'=-K}^{K} \gamma_s \gamma_{s'}\, c' M(s'-s) c$$

gilt. Nach einem bekannten Argument (s. [Malinvaud] I, S.212) ist daher z_{KT} asymptotisch normalverteilt mit Mittelwert 0 und Varianz $\sum_{s,s'=-K}^{K} \gamma_s \gamma_{s'}\, c' M(s'-s) c$.

$$3^{o}.\ \sum_{s,s'=-K}^{K} \gamma_s \gamma_{s'}\, c' M(s'-s) c = c' \sum_{s,s'=-K}^{K} \gamma_s \gamma_{s'} \int_{-\pi}^{\pi} e^{i\lambda(s'-s)}\, d\mathcal{F}\, c$$

$$= c' \int_{-\pi}^{\pi} \Big| \sum_{s=-K}^{K} \gamma_s e^{i\lambda s} \Big|^2 d\mathcal{F}\, c \xrightarrow[K\to\infty]{} 2\pi\, c' \int_{-\pi}^{\pi} f(\lambda)\, d\mathcal{F}\, c,$$

da die gleichmäßig beschränkte Folge $\Big| \sum_{s=-K}^{K} \gamma_s e^{i\lambda s} \Big|^2$ gegen $2\pi f(\lambda)$ konvergiert $(-\pi \le \lambda \le \pi)$.[1]

Folgt u_t einem stationären autoregressiven Prozeß $\sum_{r=0}^{p} \delta_r u_{t-r} = \varepsilon_t$ $(\delta_o = 1)$ mit unabhängig und identisch verteilten ε_t $(\sigma_\varepsilon^2 > 0)$, dann ist es in jedem Falle (s. etwa [Anderson], Sect. 5.2.1.) - sei der Prozeß nun invertibel oder nicht, von der Gestalt $u_t = \sum_{s=-\infty}^{\infty} \gamma_s \varepsilon_{t-s}$ mit $\sum_{s=-\infty}^{\infty} |\gamma_s| < \infty$. Also hat $K_T(\hat{\beta}-\beta)$ nach Theorem 2.5. eine asymptotische Normalverteilung mit Mittelwert 0 und der angegebenen Varianz. Für die Aitken-Schätzung β^* kann man nun zumindest bei Invertibilität des u_t-Prozesses die asymptotische Normalität von $K(\beta^*-\beta)$ ohne Schwierigkeiten zeigen:

---

1) Sofern nur $\int_{-\pi}^{\pi} f(\lambda)\, d\mathcal{F}$ regulär ist, ergibt sich unter 2^{o} für $K \ge K_o$ somit tatsächlich eine Normalverteilung.

Theorem 2.6. Folgt u_t dem invertiblen Prozeß $\sum_{r=0}^{p} \delta_r u_{t-r} = \varepsilon_t$ ($\delta_0=1$) mit unabhängig und identisch verteilten ε_t, so gilt unter der Voraussetzung A.2. für die Aitken-Schätzung β^*:
$K_T(\beta^*-\beta)$ konvergiert (bei gegebener Realisation des x_t-Prozesses) in Verteilung gegen eine normalverteilte Variable mit Mittelwert 0 und Varianz

$$\sigma_\varepsilon^2 \left(\sum_{r,s=0}^{p} \delta_r \delta_r \, M(s-r) \right)^{-1} = 2\pi \left(\int_{-\pi}^{\pi} \frac{1}{f(\lambda)} \, d\mathcal{F}(\lambda) \right)^{-1}.$$

Beweis: Nach Theorem 2.2. und (2.11) gilt

$$\lim_{T\to\infty} K_T^{-1} X' \Sigma^{-1} X K_T^{-1} = \frac{1}{\sigma_\varepsilon^2} \sum_{r,s=0}^{p} \delta_r \delta_s \, M(s-r) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{f(\lambda)} \, d\mathcal{F}(\lambda).$$

Weiterhin ist (s.S.13 f.)

$$(2.20)\quad K_T^{-1} X' \Sigma^{-1} u = K_T^{-1} X'_{(1)} \Sigma_{11}^{-1} u_{(1)} + \frac{1}{\sigma_\varepsilon^2} K_T^{-1} \sum_{t=1}^{T-p} x_{\Delta t} \, \varepsilon_{p+t}$$

mit $X'_{(1)} = [x_1, \ldots, x_p]$, $u'_{(1)} = (u_1, \ldots, u_p)$, $\Sigma_{11} = (\sigma(t-s))_{t,s=1,\ldots p}$

und $x_{\Delta t} = \sum_{r=0}^{p} \delta_r x_{p-r+t}$. Man hat

$$\max_{t\le T} | c' K_T^{-1} x_{\Delta t} | \le |c| \max_{r\le p} |\delta_r| \max_{t\le T} |K_T^{-1} x_t| \longrightarrow 0 \quad \text{(nach (2.3))} \text{ und}$$

$$\sum_{t=1}^{T-p} (c' K_T^{-1} x_{\Delta t})^2 = c' K_T^{-1} \sum_{t=1}^{T-p} \sum_{r,s=0}^{p} \delta_r \delta_s x_{t+p-r} x'_{t+p-s} K_T^{-1} c \longrightarrow c' \sum_{r,s=0}^{p} \delta_r \delta_s \, M(s-r) \, c$$

Ferner konvergiert $K_T^{-1} X'_{(1)} \Sigma_{11}^{-1} u_{(1)}$ in Wahrscheinlichkeit gegen null, woraus die Behauptung folgt. -

Unter der Voraussetzung des Theorems 2.6. gilt offensichtlich auch für die in 1.2.3. behandelte Schätzung β_Δ:

$K_T(\beta_\Delta-\beta)$ konvergiert (bei gegebener Realisation des x_t-Prozesses) in Verteilung gegen eine normalverteilte Variable mit Mittelwert 0 und derselben asymptotischen Varianz wie die Aitken-Schätzung.[1)]

---

1) Diese ist auch der Limes von $K_T V(\beta_\Delta) K_T$.

Theorem 2.7.[1] Sind unter der Vorraussetzung des Theorems 2.6. $\hat{\sigma}(0)$, $\hat{\sigma}(1),\ldots,\hat{\sigma}(p)$ konsistente Schätzungen von $\sigma(0)$ bzw. $\ldots\sigma(p)$, so setze man bei $\hat{\Sigma}_{11} := (\hat{\sigma}(t-s))_{t,s=1,\ldots,p}$

$$(2.21)\quad \begin{pmatrix}\hat{\delta}_1\\ \vdots\\ \hat{\delta}_p\end{pmatrix} = -\hat{\Sigma}_{11}^{-1}\begin{pmatrix}\hat{\sigma}(1)\\ \vdots\\ \hat{\sigma}(p)\end{pmatrix}\quad ;\ \hat{\sigma}_\varepsilon^2 = \hat{\sigma}(0) + \sum_{r=1}^{p}\hat{\delta}_r\,\hat{\sigma}(r)$$

und bestimme durch Einsetzen von $\hat{\sigma}_\varepsilon^2$ für σ_ε^2, von $\hat{\Sigma}_{11}^{-1}$ für Σ_{11}^{-1} und von $\hat{\delta}_r$ für δ_r $(r=1,\ldots,p)$ in (1.1.24) $\hat{\Sigma}^{-1}$. Für

$$(1.94)\quad \hat{\beta} = (X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}y \qquad \text{gilt dann:}$$

$K_T(\tilde{\beta}-\beta)$ konvergiert (bei gegebener Realisation des x_t-Prozesses) gegen eine normalverteilte Variable mit Mittelwert 0 und Varianz $\frac{1}{\sigma_\varepsilon^2}\sum_{r,s=0}^{p}\delta_r\delta_s M(s-r) = 2\pi\left(\int_{-\pi}^{\pi}\frac{1}{f(\lambda)}\,d\mathcal{F}(\lambda)\right)^{-1}$.

Beweis: Nach (1.97) und (1.98) sind $\hat{\sigma}_\varepsilon^2$ und $\hat{\delta}_r$ konsistente Schätzungen von σ_ε^2 bzw. δ_r $(r=1,\ldots,p)$.- Σ_{11} wird durch $\hat{\Sigma}_{11}$ konsistent geschätzt; daher ist für hinreichend großes T mit einer Wahrscheinlichkeit beliebig nahe bei 1 $\hat{\Sigma}_{11}$ regulär, $\hat{\sigma}_\varepsilon^2$ positiv und die angegebene Schätzung $\hat{\Sigma}^{-1}$ möglich. Es gilt $\operatorname{plim}_{T\to\infty} K_T^{-1}X'_{(1)}\hat{\Sigma}_{11}^{-1}X_{(1)}K_T^{-1} = \lim_{T\to\infty} K_T^{-1}X'_{(1)}\Sigma_{11}^{-1}X_{(1)}K_T^{-1} = 0$. Bezeichnet $\hat{X}_\Delta$ die $((T-p)\times n)$-Matrix mit den Zeilen $\hat{x}_{\Delta\,t} = \sum_{r=0}^{p}\hat{\delta}_r\,x_{p-r+t}$, dann gilt

$$\operatorname{plim}_{T\to\infty} K_T^{-1}\hat{X}'_\Delta\hat{X}_\Delta K_T^{-1} = \operatorname{plim}_{T\to\infty}\sum_{r,s=0}^{p}\hat{\delta}_r\hat{\delta}_s K_T^{-1}\sum_{t=1}^{T-p}x_{p-r+t}x'_{p-s+t}K_T^{-1} = \sum_{r,s=0}^{p}\delta_r\delta_s M(s-r).$$

Nun ist aber (s.(1.25)) $X'\hat{\Sigma}^{-1}X = X'_{(1)}\hat{\Sigma}_{11}^{-1}X_{(1)} + (\hat{\sigma}_\varepsilon^2)^{-1}\hat{X}'_\Delta\hat{X}_\Delta$ und also

$$(2.22)\quad \operatorname{plim}_{T\to\infty} K_T^{-1}X'\hat{\Sigma}^{-1}XK_T^{-1} = \frac{1}{\sigma_\varepsilon^2}\sum_{r,s=0}^{p}\delta_r\delta_s M(s-r).$$

Damit ist auch mit einer Wahrscheinlichkeit beliebig nahe bei 1 für hinreichend großes T $X'\hat{\Sigma}^{-1}X$ regulär und die Schätzung $\tilde{\beta}$

---

1) Man vgl. zu diesem Theorem [Hannan und Nicholls]

definiert. Es ist $K_T(\tilde{\beta}-\beta) = K_T(X'\hat{\Sigma}^{-1}X)^{-1}K_T \; K_T^{-1}X'\hat{\Sigma}^{-1}u$ und unter Berücksichtigung von (2.22) und Theorem 2.6. die Behauptung jetzt gezeigt, wenn

$$(2.23)\quad \operatorname*{plim}_{T\to\infty} K_T^{-1}X'(\hat{\Sigma}^{-1} - \Sigma^{-1})u = 0.$$

ist. Nun ist

$$K_T^{-1}X'\hat{\Sigma}^{-1}u = K_T^{-1}X'_{(1)}\hat{\Sigma}_{11}^{-1}u_{(1)} + \frac{1}{\hat{\sigma}_\varepsilon^2}K_T^{-1}\hat{X}'\hat{\varepsilon}_{(2)} \quad \text{mit}$$

$$\hat{\varepsilon}_{(2)t} = \sum_{r=0}^{p}\hat{\delta}_r u_{p-r+t} = \varepsilon_{p+t} + \sum_{r=1}^{p}(\hat{\delta}_r-\delta_r)u_{p-r+t} \qquad (t=1,\ldots,T-p).$$

Man hat daher

$$(2.24)\quad K_T^{-1}X'(\hat{\Sigma}^{-1} - \Sigma^{-1})u$$

$$= K_T^{-1}X'_{(1)}(\hat{\Sigma}_{11}^{-1}-\Sigma_{11}^{-1})u_{(1)} + \frac{1}{\delta_\varepsilon^2\sigma_\varepsilon^2}K_T^{-1}(\sigma_\varepsilon^2\sum_{t=1}^{T-p}\hat{x}_{\Delta t}\hat{\varepsilon}_{(2)t} - \hat{\sigma}_\varepsilon^2\sum_{t=1}^{T-p}x_{\Delta t}\varepsilon_{p+t}).$$

Darin ist

$$(2.25)\quad |K_T^{-1}X'_{(1)}(\hat{\Sigma}_{11}^{-1} - \Sigma_{11}^{-1})u_{(1)}| =$$

$$= |K_T^{-1}\sum_{t,s=1}^{p}(\hat{\sigma}^{ts}-\sigma^{ts})x_t u_s| \le \max_{t,s\le p}|\hat{\sigma}^{ts}-\sigma^{ts}| \max_{t\le p}|K_T^{-1}x_t| \max_{s\le p}|u_s|$$

wenn σ^{ts} bzw. $\hat{\sigma}^{ts}$ die Elemente von Σ_{11}^{-1} bzw. $\hat{\Sigma}_{11}^{-1}$ sind.

Ferner ist

$$(2.26)\quad \sigma_\varepsilon^2\sum_{t=1}^{T-p}\hat{x}_{\Delta t}\hat{\varepsilon}_{(2)t} - \hat{\sigma}_\varepsilon^2\sum_{t=1}^{T-p}x_{\Delta t}\varepsilon_{p+t} =$$

$$= \sigma_\varepsilon^2\sum_{t=1}^{T-p}\hat{x}_{\Delta t}\varepsilon_{p+t} + \sigma_\varepsilon^2\sum_{t=1}^{T-p}\hat{x}_{\Delta t}\sum_{r=1}^{p}(\hat{\delta}_r-\delta_r)u_{p-r+t} - \hat{\sigma}_\varepsilon^2\sum_{t=1}^{T-p}x_{\Delta t}\varepsilon_{p+t}$$

$$= \sigma_\varepsilon^2\sum_{t=1}^{T-p}(\hat{x}_{\Delta t}-x_{\Delta t})\varepsilon_{p+t} + (\sigma_\varepsilon^2-\hat{\sigma}_\varepsilon^2)\sum_{t=1}^{T-p}x_{\Delta t}\varepsilon_{p+t} + \sigma_\varepsilon^2\sum_{t=1}^{T-p}\hat{x}_{\Delta t}\sum_{r=1}^{p}(\hat{\delta}_r-\delta_r)u_{p-r+t}.$$

Nun gilt

$$(2.27)\quad K_T^{-1}\sum_{t=1}^{T-p}(\hat{x}_{\Delta t}-x_{\Delta t})\varepsilon_{p+t} =$$

$$= K_T^{-1}\sum_{t=1}^{T-p}\varepsilon_{p+t}\sum_{r=1}^{p}(\hat{\delta}_r-\delta_r)x_{p-r+t} = \sum_{r=1}^{p}(\hat{\delta}_r-\delta_r)\,K_T^{-1}\sum_{t=1}^{T-p}x_{p-r+t}\varepsilon_{p+t}\,.$$

Für jedes r=1,...,p konvergiert $K_T^{-1}\sum_{t=1}^{T-p}x_{p-r+t}\varepsilon_{p+t}$ in Verteilung (nämlich gegen eine nach $N(0,\sigma_\varepsilon^2\,M(0))$ verteilte Variable); deshalb gilt

$$(2.28)\quad \operatorname{plim}_{T\to\infty} K_T^{-1}\sum_{t=1}^{T-p}(\hat{x}_{\Delta t}-x_{\Delta t})\varepsilon_{p+t} = 0.$$

$K_T^{-1}\sum_{t=1}^{T-p}x_{\Delta t}\varepsilon_{p+t}$ konvergiert in Verteilung (nämlich gegen eine nach $N(0,\sigma_\varepsilon^2\sum_{r,s=0}^{p}\delta_r\delta_s\,M(s-r))$ verteilte Variable); deshalb gilt

$$(2.29)\quad \operatorname{plim}_{T\to\infty}(\sigma_\varepsilon^2-\hat{\sigma}_\varepsilon^2)\,K_T^{-1}\sum_{t=1}^{T-p}x_{\Delta t}\varepsilon_{p+t} = 0.$$

Schließlich ist

$$(2.30)\quad \sum_{t=1}^{T-p}\hat{x}_{\Delta t}\sum_{r=1}^{p}(\hat{\delta}_r-\delta_r)u_{p-r+t} =$$

$$= \sum_{r=1}^{p}(\hat{\delta}_r-\delta_r)\sum_{t=1}^{T-p}u_{p-r+t}\,(x_{p+t}+\sum_{s=1}^{p}\hat{\delta}_s x_{p-s+t})$$

$$= \sum_{r=1}^{p}(\hat{\delta}_r-\delta_r)\sum_{t=1}^{T-p}x_{p+t}u_{p-r+t} + \sum_{r,s=1}^{p}(\hat{\delta}_r-\delta_r)\hat{\delta}_s\sum_{t=1}^{T-p}x_{p-s+t}u_{p-r+t}.$$

Aufgrund von Theorem 2.5. konvergiert aber $K_T^{-1}\sum_{t=1}^{T-p}x_{p-s+t}u_{p-r+t}$ in Verteilung (nämlich gegen eine nach $N(0,2\pi\int_{-\pi}^{\pi}f(\lambda)\,d\mathcal{F}(\lambda))$ verteilte Variable); deshalb gilt

$$(2.31)\quad \operatorname{plim}_{T\to\infty} K_T^{-1}\sum_{t=1}^{T-p}\hat{x}_{\Delta t}\sum_{r=1}^{p}(\hat{\delta}_r-\delta_r)u_{p-r+t} = 0.$$

Aus (2.24), (2.25), (2.26), (2.28), (2.29) und (2.31) folgt nun (2.23). -

Für invertible Störvariablen-Prozesse gewinnt man also über die angegebene Schätzung von Σ^{-1} eine im Lichte des Theorems 2.6. asymptotisch effiziente Schätzung von β. [1)]

Voraussetzung ist allein, daß konsistente Schätzungen von $\sigma(0), \sigma(1),\ldots,\sigma(p)$ zur Verfügung stehen. Solche werden aber beispielsweise durch das Schätzverfahren (1.96) geliefert.[2)] Es genügt aber auch, über konsistente Schätzungen von $\delta_1,\ldots,\delta_p$ zu verfügen. Denn die Elemente von $\sigma_\varepsilon^2\,\Sigma_{11}^{-1}$ lassen sich aus $\delta_1,\ldots,\delta_p$ (in stetiger Weise) berechnen.[3)] Deshalb gewinnt man durch Einsetzen (gemäß (1.24)) dieser konsistenten Schätzungen anstelle von $\delta_1,\ldots,\delta_p$ in $\sigma_\varepsilon^2\,\Sigma^{-1}$ ein Schätzverfahren für β mit denselben asymptotischen Eigenschaften wie das oben angeführte. Konsistente Schätzungen für $\delta_1,\ldots,\delta_p$ kann man beispielsweise über (1.95) erlangen.[2)]

Für den Prozeß $u_t = \varrho u_{t-1} + \varepsilon_t$ $(|\varrho|<1)$ etwa (der bei unabhängig und identisch verteilten ε_t die Voraussetzungen des Theorems 2.7. erfüllt) bekommt man durch

$$(2.32)\quad \hat{\varrho} = \sum_{t=1}^{T-1} \hat{u}_{t+1}\,\hat{u}_t \Big/ \sum_{t=1}^{T-1} \hat{u}_t^2$$

in $\tilde{\beta} = (X'HX)^{-1}X'Hy$ mit

$$(2.33)\quad \begin{cases} h_{tt} = 1+\hat{\varrho}^2 \quad (t=2,\ldots,T-1);\ h_{t,t+1}=h_{t+1,t} = -\hat{\varrho} \quad (t=1,\ldots,T-1); \\ h_{11} = h_{TT} = 1;\quad h_{tt'} = 0 \quad \text{für} \quad 2\leq|t-t'|\leq T-1 \end{cases}$$

einen asymptotisch normalverteilten und effizienten Schätzer von β.

---

1) Natürlich nicht nach dem in 2.3. gebrauchten Effizienzbegriff, sondern in dem Sinne, daß ihre asymptotische Verteilung mit derjenigen der Aitken-Schätzung übereinstimmt.

2) s. 2.5.

3) s. [Anderson], S.576, (85) in Zusammenhang mit (1.24).

Als Spezialfall enthält Theorem 2.5. übrigens die asymptotische Normalverteilung von $K_T(\hat{\beta}-\beta)$ im Fall unabhängig und identisch verteilter u_t. Die asymptotische Varianz ist dann gerade $\sigma^2 M^{-1}(0)$.

2.5. Konsistenz der Parameter-Schätzungen

Bei unkorrelierten (homoskedastischen) u_t ist die Kleinstquadrat-Schätzung $\hat{\beta}$ von β konsistent, sofern nur $\lim_{T\to\infty} (X'X)^{-1} = 0$ ist. Das ist beispielsweise immer dann der Fall, wenn

$\sum_{t=1}^{\infty} x_{it}^2 = \infty$ ist, $\lim_{T\to\infty} \sum_{t=1}^{T} x_{it}x_{jt} / \sqrt{\sum_{t=1}^{T} x_{it}^2 \sum_{t=1}^{T} x_{jt}^2} = m_{ij}(0)$ existiert

$(i,j=1,\ldots,n)$ und $M(0) = (m_{ij}(0))_{i,j=1,\ldots,n}$ regulär ist.

Auch bei autokorrelierten u_t ist $\hat{\beta}$ eine konsistente Schätzung, wenn die Voraussetzungen A.1.(i) und A.2.(i) erfüllt sind. Das ist eine unmittelbare Folge von Theorem 2.1.. Tatsächlich ist diese Voraussetzung aber hinsichtlich (der Realisation) des x_t-Prozesses zu stark gefaßt, denn es gilt:

Theorem 2.8. Unter der Voraussetzung A.1.(i) sind Aitken- und Kleinstquadratschätzung (bei gegebener Realisation des x_t-Prozesses) konsistente Schätzungen von β, wenn $\lim_{T\to\infty} (X'X)^{-1} = 0$ ist.

Beweis: Nach (1.50) ist $V(\beta^*) \leq V(\hat{\beta}) \leq 2\pi \max f(\lambda)\ (X'X)^{-1}$. -

Andererseits läßt sich die Konsistenz der Kleinstquadrat-Schätzung (und damit der Aitken-Schätzung) selbst dann sichern, wenn man über die Struktur des u_t-Prozesses außer der Stationaritätsannahme überhaupt keine Voraussetzung macht. Die Forderung an die Struktur (der Realisation) des x_t-Prozesses muß nur genügend verschärft werden. Beispielsweise genügt die Voraussetzung des Theorems 2.3..

Auch der Parameter σ^2 wird im klassischen Modell nach der Kleinstquadrat-Methode konsistent geschätzt. Ist $Eu_t^4 < \infty$ und sind die u_t unabhängig und identisch verteilte Variablen (oder verhalten sich in ihren Momenten bis zur Ordnung 4 wie solche), dann entnimmt man 4.3., daß $\hat{\sigma}^2$ von (1.7) sogar (ebenso wie die $\hat{\beta}_i$ bei $(X'X)^{-1} \to 0$) im quadratischen Mittel gegen σ^2 konvergiert. Denn es ist

$$(2.34)\quad V(\hat{\sigma}^2|X) = \frac{2\sigma^4}{T-n} + \frac{1}{(T-n)^2} \sum_{t=1}^{T} p_{tt}^2 \,(Eu^4 - 3\sigma^4)$$

mit den Diagonalelementen p_{tt} von $P = I - X(X'X)^{-1}X'$, wobei $0 \leq p_{tt} \leq 1$ ist $(t=1,\ldots,T)$. Ist $Eu^4 = \infty$, so ist $\hat{\sigma}^2$ jedenfalls noch konsistent, wenn $K_T^{-1}X'XK_T^{-1}$ für eine Folge von Diagonalmatrizen K_T gegen eine reguläre Matrix konvergiert, wenn die $k_i(T)$ $(i=1,\ldots,n)$ $k_i(T) \to \infty$ und $k_i(T-1)/k_i(T) \to 1$ erfüllen.[1)]

Allgemein ist unter den Voraussetzungen A.2. und A.1.' bei $E\varepsilon_t^4 < \infty$

$$(2.35)\quad \hat{\sigma}(\tau) = \frac{1}{T-\tau} \sum_{t=1}^{T-\tau} \hat{u}_{t+\tau}\, \hat{u}_t \qquad (\tau \geq 0)$$

eine konsistente Schätzung von $\sigma(\tau)$. Es gilt nämlich

$\hat{u}_t = u_t - x_t' \,(X'X)^{-1}X'u$ und daher

$$\hat{\sigma}\,(\tau) = \frac{1}{T-\tau}\Big[\sum_{t=1}^{T-\tau} u_{t+\tau}u_t - u'X(X'X)^{-1} \sum_{t=1}^{T-\tau} x_t u_{t+\tau} - u'X(X'X)^{-1} \sum_{t=1}^{T-\tau} x_{t+\tau}u_t$$
$$+ u'X(X'X)^{-1} \sum_{t=1}^{T-\tau} x_{t+\tau}x_t' \,(X'X)^{-1}X'u\Big].$$

Nach Theorem 2.5. konvergieren $K_T^{-1}X'u$, $K_T^{-1}\sum_{t=1}^{T-\tau} x_{t+\tau}u_t$ und $K^{-1}\sum_{t=1}^{T-\tau} x_t u_{t+\tau}$

---

1) Um dies zu zeigen, geht man analog vor wie im folgenden Beweis. Man hat plim $K_T^{-1}X'u/\sqrt{T} = 0$ und für die identisch und unabhängig verteilten u_t: plim $\sum u_t^2/T = \sigma^2$.

jeweils in Verteilung; Division durch $\sqrt{T}$ liefert also jeweils eine Variable mit Wahrscheinlichkeitslimes null. Deshalb ist (da $K_T(X'X)^{-1}K_T$ konvergiert)

$$\operatorname*{plim}_{T\to\infty} (\hat{\sigma}(\tau) - \frac{1}{T-\tau} \sum_{t=1}^{T-\tau} u_{t+\tau}\, u_t) = 0$$

Ferner ist der Prozeß A.1.'kovarianzergodisch, also

$$\operatorname*{plim}_{T\to\infty} \frac{1}{T-\tau} \sum_{t=1}^{T-\tau} u_{t+\tau}\, u_t = \sigma(\tau),$$

woraus die Behauptung folgt.[1)]

Schließlich werde noch die Konsistenz der Schätzungen $\hat{\delta}_1,\ldots,\hat{\delta}_p$ nach (1.95) gezeigt. $\hat{\delta}_1,\ldots,\hat{\delta}_p$ sind die Lösungen des Systems

$$\hat{u}_t \doteq - \sum_{r=1}^{p} \hat{\delta}_r\, \hat{u}_{t-r} \qquad (t=p+1,\ldots,T).$$

Benennt man $\hat{u}'_{-r} := (\hat{u}_{p+1-r},\ldots,\hat{u}_{T-r}) = (\hat{u}_{p-r+t})'_{t=1,\ldots,T-p}$ $(r=0,1,\ldots,p)$ und $\hat{U} := [\hat{u}_{-1},\ldots,\hat{u}_{-p}]$, $\hat{\delta}' := (\hat{\delta}_1,\ldots,\hat{\delta}_p)$, dann ist

$$(2.36)\quad \hat{\delta} = - (\hat{U}'\hat{U})^{-1}\hat{U}'\hat{u}_o.$$

Es gilt

$$\hat{u}'_{-r}\, \hat{u}_{-s} = \sum_{t=1}^{T-p} \hat{u}_{p-r+t}\, \hat{u}_{p-s+t}.$$

Setzt man hierin $\hat{u}_h = u_h - x'_h(X'X)^{-1}X'u$ ein, dann läßt sich wie oben zeigen:

Erfüllt u_t A.1.' nebst $E\epsilon_t^4 < \infty$ und x_t A.2., dann gilt

$$\operatorname*{plim}_{T\to\infty} \hat{u}'_{-r}\hat{u}_{-s}/T = \operatorname*{plim}_{T\to\infty} \sum_{t=1}^{T-p} u_{p-r+t}u_{p-s+t}/T.$$

---

1) Tatsächlich ist sogar $\sqrt{T}(\hat{\sigma}(\tau) - \sigma(\tau))$ asymptotisch normalverteilt ([Anderson], Theorem 8.4.2. und 10.3.3.).

$\sum_{t=1}^{T-p} u_{p-r+t}u_{p-s+t}/T$ weicht von $\sum_{t=1}^{T-|r-s|} u_{t+|r-s|}u_t/T$ lediglich um $p-|r-s|$ Summanden ab und konvergiert im quadratischen Mittel gegen $\sigma(|r-s|)$. Damit ist $-\Sigma_{11}^{-1}(\sigma(r))_{r=1,\ldots,p}$ der Wahrscheinlichkeitslimes der rechten Seite von (2.36). Wird deshalb u_t von dem invertiblen autoregressiven Prozeß $\sum_{r=0}^{p}\delta_r u_{t-r} = \varepsilon_t \quad (\delta_o=1)$ hervorgebracht, dann folgt aus (1.98):

$$(2.37)\quad \operatorname*{plim}_{T\to\infty} \hat{\delta}_r = \delta_r \qquad (r=1,\ldots,p)$$

Benutzt man unter diesen Voraussetzungen die obigen Schätzungen $\hat{\sigma}(\tau)$ $(\tau=0,\ldots,p)$ in der in Theorem 2.7. angezeigten Weise oder diese Schätzungen $\hat{\delta}_r$ $(r=1,\ldots,p)$ in der S.86 angezeigten Weise zur Konstruktion einer approximativen Aitken-Schätzung, dann ist dieselbe asymptotisch normalverteilt, jedenfalls aber auch konsistent und in der linearen Schätzklasse asymptotisch effizient.

2.6. Trendbehaftete Regressoren

2.6.1. Polynomialer Trend

Die Regressoren mögen polynomialen Trends folgen:

$$(2.38)\quad x_{it} = q_{im_i}t^{m_i} + \ldots + q_{i1}t + q_{io} \qquad (q_{im_i} \neq 0)$$

Insbesondere ist auch die Scheinvariable, falls eine solche vorkommt, von dieser Gestalt. Es gilt

$$(2.39)\quad \sum_{t=1}^{T-\tau} x_{it+\tau}x_{jt} = \sum_{t=1}^{T-\tau} x_{it}x_{jt} + \sum_{t=1}^{T-\tau}(x_{it+\tau} - x_{it})x_{jt}$$

$$= \frac{q_{im_i}q_{jm_j}}{m_i+m_j+1}(T-\tau)^{m_i+m_j+1} + O(T^{m_i+m_j}).\ ^{1)}$$

---

1) Vgl. die Betrachtung S.28 f.

Offensichtlich sind mit

$$(2.40)\quad k_i(T) = T^{m_i+\frac{1}{2}} \qquad (i=1,\ldots,n)$$

die Voraussetzungen der Annahme A.2. erfüllt, und es liegt der spezielle Fall vor, daß

$$(2.41)\quad M(\tau) \equiv M(0) \qquad (\tau=0,\pm1,\ldots)$$

ist. M(0) ist genau dann regulär, wenn die m_i (i=1,...,n) paarweise verschieden sind, d.h., wenn die Polynome lauter verschiedene Grade haben:
Denn es gilt

$$(2.42)\quad M(0) = Q\left[\frac{1}{m_i+m_j+1}\right]_{i,j=1,\ldots,n}Q$$

mit $Q = \mathrm{diag}(q_1,\ldots,q_n)$. Zieht man in $\det\left[\frac{1}{m_i+m_j+1}\right]$ von jeder Zeile i=1,...,n-1 das $\{(2m_n+1)/(m_i+m_n+1)\}$ -fache von Zeile n ab, dann ergibt sich

$$(2.43)\quad \det\left[\frac{1}{m_i+m_j+1}\right]_{i,j=1,\ldots,n} = \prod_{k=1}^{n-1}\left(\frac{m_n-m_k}{m_n+m_k+1}\right)^2 \det\left[\frac{1}{m_i+m_j+1}\right]_{i,j=1,\ldots,n-1}$$

Durch Induktion folgt die Behauptung.

Sind die m_i also paarweise verschieden, dann ist wegen (2.41) nach Theorem 2.4. die Kleinstquadratschätzung asymptotisch effizient für jeden Störvariablen-Prozeß mit stetiger positiver Spektraldichte. Ihre asymptotische Varianz ist $2\pi f(0)\, M^{-1}(0)$.[1)]

Für den autoregressiven Prozeß erster Ordnung ist diese Matrix gleich $\sigma^2\frac{1+\varrho}{1-\varrho}\, M^{-1}(0)$, was bei ϱ nahe 1 eine beträchtliche Vergrößerung der asymptotischen Varianz gegenüber dem aurokorrelationsfreien Modell bedeutet.[2)]

---

1) Dieses Ergebnis findet sich auch bei [Grenander und Rosenblatt], S.247. - Nach (2.41) hat $\mathfrak{F}$ nur eine Wachstumsstelle, nämlich $\lambda=0$.

2) Man vgl. (1.54)

Zu genau den gleichen Aussagen kommt man, wenn man - etwas realistischer - zuläßt, daß die (eigentlichen) Regressoren nicht exakt, sondern zufallsgestört trendabhängig sind. Fügt man in (2.38) eine Störung w_{it} ein, derart, daß der Prozeß $\{w_t\}$ mit $w_t' = (w_{1t},\ldots,w_{nt})$ ein reiner Zufallsprozeß (d.h. unabhängig und identisch verteilt mit $Ew_t = 0$, $Ew_t w_s' = \delta_{ts}\Sigma_w$) ist, dann ergibt sich genau wie oben $M(\tau) \equiv M(0)$, sofern nur der Polynomgrad m_i (ausgenommen allenfalls die Scheinvariable) positiv ist.[1)]

2.6.2. Exponentieller Trend

Nunmehr möge

$$(2.44)\quad x_{jt} = a_j e^{q_j t} \qquad (a_j, q_j > 0;\ j=1,\ldots,n)$$

gelten.[2)] Man erhält für $\tau \geq 0$

$$(2.45)\quad \sum_{t=1}^{T-\tau} x_{jt+\tau}x_{lt} = a_j a_l \sum_{t=1}^{T-\tau} e^{(q_j+q_l)t}\, e^{q_j\tau}$$

Mit

$$(2.46)\quad k_j(T) = e^{q_j T}$$

ergibt sich

---

1) Man bekommt mit $k_i(T) = T^{m_i+\frac{1}{2}}$ $(i=1,\ldots,n)$ $K_T^{-1}N^T(\tau)K_T^{-1} \to M(0)$ mit Wahrscheinlichkeit 1. - Im Beweis wird wesentlich genutzt, daß für jede Zahl $m \geq 0$ $\sum_{t=1}^{T} t^m w_t / T^{m+1}$ mit Wahrscheinlichkeit 1 gegen null geht. Da die w_t's unabhängig und identisch verteilt sind, genügt es hierfür - nach einem Argument von [Loève], S.238, - zu zeigen, daß $\sum_{t=1}^{T} \frac{t^{2m}}{t^{2(m+1)}} < \infty$ ist, was aber der Fall ist.

2) Ebenso wie für den Polynomtrend darf man hier eine reine Zufallsüberlagerung des Trendterms zulassen, ohne daß das nachfolgende Ergebnis beeinträchtigt würde.

$$\sum_{t=1}^{T-\tau} x_{jt+\tau}x_{1t}/k_j(T)k_1(T) =$$

$$= a_j a_1 e^{q_j\tau} e^{q_j+q_1} (e^{(q_j+q_1)(T-\tau)}-1)/(e^{q_j+q_1}-1)\, e^{(q_j+q_1)T} =$$

$$= e^{-q_1\tau} a_j a_1 e^{q_j+q_1} \{1 - e^{-(q_j+q_1)(T-\tau)}\}/(e^{q_j+q_1}-1).$$

Man setze für $-\pi\leq\lambda\leq\pi$

$$(2.47)\quad f_{j1}^T(\lambda) = \frac{1}{2\pi k_j(T)k_1(T)}\left[\sum_{t=1}^{T} x_{jt}x_{1t} + \right.$$

$$\left. + \sum_{\tau=1}^{T-1}\{e^{-i\lambda\tau}\sum_{t=1}^{T-\tau} x_{j,t+\tau}x_{1t} + e^{i\lambda\tau}\sum_{t=1}^{T-\tau} x_{jt}x_{1t+\tau}\}\right] =$$

$$= \frac{a_j a_1 e^{q_j+q_1}}{2\pi(e^{q_j+q_1}-1)}\left[(1 - e^{-(q_j+q_1)T}) + \right.$$

$$\left. + \sum_{\tau=1}^{T-1}\{e^{-(q_1+i\lambda)\tau} + e^{-(q_j-i\lambda)\tau}\}\{1 - e^{-(q_j+q_1)(T-\tau)}\}\right].$$

$\sum_{\tau=1}^{T-1}|e^{-(q_1+i\lambda)\tau} + e^{-(q_j-i\lambda)\tau}|$ ist gleichmäßig beschränkt durch die konvergente Reihe $\sum_{\tau=1}^{\infty}\{e^{-q_1\tau} + e^{-q_j\tau}\}$ und $1 - e^{-(q_j-q_1)(T-\tau)}$ konvergiert für festes τ gleichmäßig beschränkt in τ gegen 1. Deshalb konvergiert $f_{j1}^T(\lambda)$ gleichmäßig beschränkt gegen

$$(2.48)\quad f_{j1}(\lambda) = \frac{a_j a_1 e^{q_j+q_1}}{2\pi(e^{q_j+q_1}-1)}\left[1 + \frac{1}{e^{q_1+i\lambda}-1} + \frac{1}{e^{q_j-i\lambda}-1}\right]$$

$$= \frac{1}{2\pi} a_j a_1 e^{q_j+q_1} \frac{1}{(e^{q_1+i\lambda}-1)(e^{q_j-i\lambda}-1)}\,.$$

$$(2.49)\quad \mathcal{f}_T(\lambda) := (f_{j1}^T(\lambda))_{j,1=1,\ldots,n}$$

erfüllt daher die Vorraussetzung des Theorems 2.3. und mit

$$(2.50)\quad f(\lambda) := \lim_{T\to\infty} f_T(\lambda)$$

gilt für jeden stationären u_t-Prozeß (mit der zugehörigen Spektralverteilungsfunktion $F(\lambda)$):

$$(2.51)\quad \lim_{T\to\infty} K_T V(\hat{\beta}) K_T = 2\pi M^{-1}(0) \int_{-\pi}^{\pi} f(\lambda)\,dF(\lambda)\, M^{-1}(0). \quad 1)$$

Für jeden stationären u_t-Prozeß ist also die Kleinstquadrat-Schätzung konsistent. Hingegen ist sie im Lichte des Theorems 2.4. für u_t- Prozesse mit positiver stetiger Spektraldichte nicht effizient. Denn wie man (2.48) entnimmt, ist $f(\lambda)$ überall stetig und verschwindet nirgends; deshalb besteht das Spektrum von $\mathcal{F}$ in diesem Fall aus dem gesamten Intervall $[-\pi,\pi]$. 2)

---

1) $m_{ij}(0)$ ist von der Form $h_i h_j/(e^{q_i+q_j}-1)$ $(i,j=1,\ldots,n;$ $h_i\neq 0,\ i=1,\ldots,n)$. Nach genau demselben Muster wie in dem entsprechenden Beweis S.91 ergibt sich: M(0) ist genau dann regulär, wenn die q_i $(i=1,\ldots,n)$ paarweise verschieden sind.

2) Im übrigen ist (nach dem Satz von der majorisierten Konvergenz) $f(\lambda)$ die Spektraldichte der Folge $\{M(\tau)\}$.

3. Zusammenfassung

In kleiner Stichprobe ist bei autokorrelierter Störvariable die Kleinstquadrat-Schätzung zwar weiterhin unverzerrte, aber in der Regel nicht länger mehr wie im autokorrelationsfreien Modell auch beste (lineare unverzerrte) Schätzung. Dies verlangt vielmehr einen sehr speziellen Zusammenhang zwischen den Spalten von X (i.e.: den Datenreihen der Regressoren) und den Eigenvektoren von Σ; oder, in äquivalenter Formulierung, eine sehr spezielle Struktur der Spektralzerlegung von X'X in bezug auf Σ (s. 1.2.2.4.).

Unter Zugrundelegung bestimmter Vorstellungen über 'die allgemein in der Zeitreihenökonometrie vorgefundene Situation'[1] lassen sich näherungsweise die folgenden gängigen Meinungen begründen:

In Zeitreihenregressionen wird nach der Kleinstquadrat-Methode die Restvarianz σ^2 systematisch unterschätzt. Aus diesem Grunde - und weil bereits auch $X'\Sigma X \geq \sigma^2 X'X$ ist - liefern auch die nach der Kleinstquadrat-Methode geschätzten Varianzen der Koeffizientenschätzungen systematische Unterschätzungen (1.2.2.2. und 1.2.2.3.).

In Unkenntnis von Σ (oder eines Vielfachen von Σ) bzw. - wenn u_t dem autoregressiven Prozeß $\sum_{r=o}^{p} \delta_r u_{t-r} = \varepsilon_t$ ($\delta_o = 1$) folgt -

---

1) nämlich der Annahme, daß die Autokovarianzfunktion $\sigma(\tau)$ der Störvariablen (zumindest für betragsmäßig kleines τ) positiv ist und ziemlich bald (für $|\tau| \geq \tau_o$ mit kleinerem τ_o) (betragsmäßig) klein (gegenüber den ersten Autokovarianzen); und der (vgl. die Fn. 1, S. 6 zitierten Stellen bei [Schulmann]) empirisch gestützten weiteren Annahme, daß die Regressoren positiv auto- und reihenkorreliert sind, sich beispielsweise durch Trendpolynome gut approximieren lassen (diese Annahme kann ganz analog wie die Annahme über die Autokovarianzfunktion $\sigma(\tau)$ und auch mit denselben Abschwächungen als Annahme über die Automomentenfunktion der Beobachtungen x_t formuliert werden).

der Parameter $\delta_1,\ldots,\delta_p$ läßt sich nun weder die beste lineare unverzerrte Schätzung, die Aitken-Schätzung nämlich, noch die (in 1.2.3. beschriebene) Kleinstquadrat-Schätzung nach prozeßkonformer Modelltransformation ausführen, deren Varianz nur geringfügig von derjenigen der Aitken-Schätzung abweicht. Man kann aber wenigstens versuchen, über eine Schätzung von Σ (beispielsweise mittels der Kleinstquadrat-Residuen) eine Approximation der Aitken-Schätzung zu erreichen; die Eigenschaften dieser approximativen Aitken-Schätzung bleiben allerdings im Dunkeln (s. 1.2.5.). In dem Sonderfall, daß u_t einem autoregressiven Prozeß erster Ordnung (mit hohem Autokorrelationskoeffizienten erster Ordnung) folgt, hat man in der Kleinstquadrat-Schätzung im Differenzen-Modell ein praktikables und gutes Schätzverfahren (s. 1.2.4.).

Gleichgültig aber, welches Schätzverfahren man wählt: Prinzipiell fällt die Koeffizientenschätzung infolge von Kollinearität (in den Spalten) der Matrix X ungenau aus [1] (s. 1.1.2., 1.2.1.3., 1.2.2.2., 1.2.2.3., 1.2.6.).

I n g r o ß e r S t i c h p r o b e ist dagegen unter allgemeinen Bedingungen die Welt wieder heil. Die Kleinstquadrat-Schätzung von β ist konsistent und auch σ^2 wie überhaupt alle $\sigma(\tau)$ (τ=0,1,...) lassen sich mittels der Kleinstquadrat-Residuen konsistent schätzen; dies erstreckt sich auch auf die Parameter $\delta_1,\ldots,\delta_p$, wenn u_t dem obigen autoregressiven Prozeß folgt (s.2.5.). In letzterem Fall verfügt man in der approximativen Aitken-Schätzung sogar über einen asymptotisch normalverteilten und effizienten Schätzer für β [2]. Dies ist

---

1) Die Ungenauigkeit ist modell-inhärent.

2) Unter der Normalverteilungsannahme für die u_t ist diese Schätzung, wie Theorem 1.7. zu entnehmen ist, nicht allein in der linearen Schätzklasse, sondern auch in der Klasse aller regulären Schätzer effizient.

(in Verbindung mit Aussagen von 2.5.) der Inhalt des Theorems 2.7. [1].

Allgemein ist auch die Kleinstquadrat-Schätzung von ß asymptotisch normalverteilt (s. 2.4.), asymptotisch effizient jedoch nur dann, wenn die Autokovarianzfunktion des u_t-Prozesses und die Automomentenfunktion (der Realisation) des x_t-Prozesses in einer sehr speziellen Beziehung stehen (s. 2.3.). Für polynomial trendbehaftete Regressoren ist die Kleinstquadrat-Schätzung (fast) stets effizient. Im allgemeinen werden die asymptotischen Varianzen nach der Kleinstquadrat-Methode unterschätzt (s. 2.2.).

Die asymptotischen Varianzen von Aitken- und Kleinstquadrat-Schätzung können häufig nur in Spektral-Repräsentation angegeben werden (s. 2.2.). Es läßt sich zeigen, daß eine völlig analoge Repräsentation (im wesentlichen) auch für die endliche Stichprobe gilt (s. 1.2.1.2., 1.2.2.1.) und analoge Konsequenzen hat (s. 1.2.2.5.).

---

1) Für Spezialfälle übrigens von [Schönfeld], S. 21o ff. und [Malinvaud] II, S. 513 ff., bewiesen.

4. Anhang

4.1. Matrizen

4.1.1. Die Inversion der unterteilten (quadratischen) regulären Matrix

$$A = \begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix}$$

vollzieht sich, wenn die quadratische Matrix A_4 regulär ist, nach der bekannten Formel

$$A^{-1} = \begin{bmatrix} A^1 & -A^1 A_2 A_4^{-1} \\ -A_4^{-1} A_3 A^1 & A_4^{-1} + A_4^{-1} A_3 A^1 A_2 A_4^{-1} \end{bmatrix} \tag{4.1.1}$$

mit $A^1 = (A_1 - A_2 A_4^{-1} A_3)^{-1}$; der Ausdruck in (.) ist regulär.

Hat die relle Matrix $Z = [Z_1, Z_2]$ vollen Spaltenrang, so gilt insbesondere für $A = Z'Z$:

$A_1 = Z_1'Z_1$, $A_2 = Z_1'Z_2 = A_3'$, $A_4 = Z_2'Z_2$ und

$$A^1 = (Z_1'P_{Z_2}Z_1)^{-1} \tag{4.1.2}$$

mit

$$P_{Z_2} := I - Z_2(Z_2'Z_2)^{-1}Z_2' \; . \tag{4.1.3}$$

A^1 ist in diesem Fall positiv-definit.

Für den entsprechenden oberen Teilvektor v_1 von $v = (Z'Z)^{-1}Z'y$ ergibt sich

$$v_1 = (A^1Z_1' - A^1Z_1'Z_2(Z_2'Z_2)^{-1}Z_2')y = (Z_1'P_{Z_2}Z_1)^{-1}Z_1'P_{Z_2}Y \; . \tag{4.1.4}$$

Ist Ω eine reelle, symmetrische positiv-definite Matrix, dann erhält man bei Anwendung des Vorstehenden auf $Z_\phi := \phi^{-1}Z$ anstelle von Z, wobei $\Omega = \phi\phi'$ gilt:

(4.1.2') $A^1 = \{Z_1'(\Omega^{-1} - \Omega^{-1}Z_2(Z_2'\Omega^{-1}Z_2)^{-1}Z_2'\Omega^{-1})Z_1\}^{-1}$

und

(4.1.4') $v_{1\phi} = A^1 Z_1'(\Omega^{-1} - \Omega^{-1}Z_2(Z_2'\Omega^{-1}Z_2)^{-1}\ Z_2'\Omega^{-1})y$.

4.1.2. Ist die (m x m)-Matrix A regulär, B m x n und C n x m, dann gilt, wie man durch Multiplikation mit A + BC bestätigt,

(4.1.5) $(A+BC)^{-1} = A^{-1} - A^{-1}B(I + CA^{-1}B)^{-1}CA^{-1}$,

sofern nur $I + CA^{-1}B$ regulär ist. Ist B ein m-Vektor b und $C = \lambda b'$ mit einem Skalar λ, so hat man

(4.1.5') $(A + \lambda bb')^{-1} = A^{-1} - \lambda(1 + \lambda b'A^{-1}b)^{-1}\ A^{-1}bb'A^{-1}$.

4.1.3. Ist A m x n und B n x m , dann haben AB und BA dieselben nichtverschwindenden Eigenwerte, und die entsprechenden Eigenräume zum gleichen Eigenwert haben gleiche Dimension:

Ist nämlich etwa $ABV = \lambda V$ mit $\lambda \neq 0$ und V von vollem Spaltenrang, so hat ABV und also erst recht BV vollen Spaltenrang und es gilt $BA(BV) = \lambda(BV)$. -

4.1.4. Ist die (m x m)-Matrix A reell, symmetrisch und definit, B reell m x n und λ eine reelle Zahl, so gilt

(4.1.6) $\det(A + \lambda BB') = \det A \ \det(I + \lambda B'A^{-1}B)$,

und insbesondere für einen reellen m-Vektor b:

(4.1.6') $\det(A + \lambda bb') = (1 + \lambda b'A^{-1}b)\det A.$

Beweis: Es sei $\lambda \neq 0$. Dann ist Rang $(\lambda B'A^{-1}B)$ = Rang $(\lambda A^{-1}BB')$ = = Rang (B).
Für die symmetrische Matrix $\lambda B'A^{-1}B$ ist für jeden Eigenwert die Vielfachheit gleich der Dimension des zugehörigen Eigenraumes. Folglich ist die Summe der Dimensionen der zu den nichtverschwindenden Eigenwerten gehörigen Eigenräume für λ $B'A^{-1}B$ und (vgl. 4.1.3.) also auch für $\lambda A^{-1}BB'$ gleich Rang (B).

Ferner gehört zu letzterer Matrix der Eigenwert 0 mit der Dimension m-Rang(B). Die Summe der Dimensionen aller Eigenräume ist also für $\lambda A^{-1}BB'$ gleich m, und folglich auch für diese Matrix die Vielfachheit jedes Eigenwertes gleich der Dimension des zugehörigen Eigenraumes. $\lambda B'A^{-1}B$ und $\lambda A^{-1}BB'$ haben somit (nach 4.1.3.) dieselben nichtverschwindenden Eigenwerte mit je gleicher Vielfachheit. Sind λ_j (j = 1,..., Rang (B)) diese Eigenwerte, so folgt

$$\det (I + \lambda B'A^{-1}B) = \Pi(1 + \lambda_j) = \det (I + \lambda A^{-1}BB') = \det A^{-1} \det(A+\lambda BB')$$

4.1.5. Es sei B eine positiv-definite hermitesche Matrix, d ein (komplexer) Vektor. Dann gilt

(4.1.7) $$\max_{c \neq 0} \frac{|c^*d|^2}{c^*Bc} = d^*B^{-1}d.$$

Beweis: Es sei $d \neq 0$. Bezeichnet man
$g(c) := |c^*d|^2 / c^*Bc = c^*dd^*c / c^*Bc$, so ist g(c) höchstens gleich dem größten Eigenwert des Problems $dd^* - \lambda B$.
Man hat also für das betreffende c: $dd^*c = \lambda Bc$ mit $\lambda \neq 0$. Es folgt $c = \mu B^{-1}d$ $(\mu \neq 0)$ und daher $g(c) = |\mu|^2(d^*B^{-1}d)^2/|\mu|^2 d^*B^{-1}d = d^*B^{-1}d$.

4.2. Summenformeln mit der Trendvariablen

4.2.1. Mittels der bekannten Formeln $\sum_1^T t = T(T+1)/2$ und $\sum_1^T t^2 = T(T+1)(2T+1)/6$ gewinnt man sofort

$$(4.2.1)\qquad \bar{t} := 1/_T \sum_1^T t = (T+1)/2 \; ; \quad s_t^2 := 1/_T \sum_1^T (t-\bar{t})^2 = (T^2-1)/12.$$

Der Übergang von $\bar{t}_{(1)} := \frac{1}{T-\tau} \sum_1^{T-\tau} t$ zu $\bar{t}_{(2)} := \frac{1}{T-\tau} \sum_{\tau+1}^{T} t$ bzw. von

$$s^2_{t(1)} := \frac{1}{T-\tau} \sum_1^{T-\tau} (t-\bar{t}_{(1)})^2 \quad \text{zu} \quad s^2_{t(2)} := \frac{1}{T-\tau} \sum_{\tau+1}^{T} (t-\bar{t}_{(2)})^2$$

entspricht offenbar einer Translation der t-Werte um den Betrag τ, so daß also unter Verwendung von (4.2.1)

$$(4.2.2)\qquad \begin{cases} \bar{t}_{(1)} = (T-\tau+1)/2 \; ; \; \bar{t}_{(2)} = \bar{t}_{(1)} + \tau = (T+\tau+1)/2 \; ; \\ s^2_{t(2)} = s^2_{t(1)} = ((T-\tau)^2 - 1)/12 \end{cases}$$

gilt. Weiter ist

$$s_{t\tau} := \sum_{\tau+1}^{T} (t-\bar{t}_{(2)})(t-\tau-\bar{t}_{(1)}) / (T-\tau) = \sum_{\tau+1}^{T} (t-\tau-\bar{t}_{(1)})^2/(T-\tau) = s^2_{t(1)},$$

so daß also sämtliche Autokorrelationskoeffizienten $r_{t\tau} = s_{t\tau}/s_{t(1)}s_{t(2)}$ den Wert 1 haben $(\tau = 1,\ldots,T-2)$.

4.2.2. Für ganzzahliges $m \geq 0$ ist

$$\sum_1^{T-1} \varrho^\tau \tau^{m+1} = \sum_1^{T-1} \sum_{\tau=t}^{T-1} \varrho^\tau \tau^m = (T-1) \sum_1^{T-1} \varrho^\tau \tau^m - \sum_{t=1}^{T-1} \sum_{\tau=1}^{t-1} \varrho^\tau \tau^m \text{ , d.h.}$$

$$(4.2.3)\qquad \sum_{t=1}^{T-1} \varrho^\tau \tau^{m+1} = (T-1) \sum_1^{T-1} \varrho^\tau \tau^m - \sum_{t=1}^{T-2} \sum_{\tau=1}^{t} \varrho^\tau \tau^m \; .$$

Man hat $\sum_1^{T-1} \varrho^\tau = \varrho(1-\varrho^{T-1}) / (1-\varrho)$. Unter Benutzung der Rekursionsformel (4.2.3) gewinnt man durch elementare Rechnung sukzessive

$$(4.2.4)\qquad \sum_1^{T-1} \varrho^\tau \tau = \varrho\{(T-1)(1-\varrho^{T-1}) - \sum_1^{T-2}(1-\varrho^t)\}/(1-\varrho)$$

$$= \varrho(1-\varrho)^{-2}\left[1-\varrho^{T-1}\{1+(T-1)(1-\varrho)\}\right] \quad ;$$

$$(4.2.5)\qquad \sum_1^{T-1} \varrho^\tau \tau^2 = \varrho(1-\varrho)^{-2}\left[(T-1)\left[1-\varrho^{T-1}\{1+(T-1)(1-\varrho)\}\right] - \sum_1^{T-2}\left[1-\varrho^t\{1+t(1-\varrho)\}\right]\right]$$

$$= \varrho(1-\varrho)^{-3}\left[1-\varrho^{T-1}\{1+(T-1)(1-\varrho)\}^2 + \varrho(1-\varrho^{T-1})\right] \quad ;$$

$$(4.2.6)\qquad \sum_1^{T-1} \varrho^\tau \tau^3 = \varrho(1-\varrho)^{-4}\left[1-\varrho^{T-1}\{1+(T-1)(1-\varrho)\}^3 - 3T\varrho^T(1-\varrho) + \right.$$

$$\left. + 4\varrho(1-\varrho^T) + \varrho^2(1-\varrho^{T-2})\right] \quad .$$

4.2.3. Anwendung von (4.2.3) auf $\varrho = 1$ liefert einen Induktionsbeweis für die bekannte Aussage

$$(4.2.7)\qquad \sum_1^T t^m = \frac{T^{m+1}}{m+1} + O(T^m) .$$

4.3. Varianz einer quadratischen Form

Für die stochastisch unabhängigen Zufallsvariablen $u_1,\ldots,u_T$ mit $E\,u_t^4 = \mu_4 (<\infty)$, $E\,u_t^2 = \sigma^2$, $E\,u_t = 0$ ergibt sich als Varianz der quadratischen Form $u'Au$ (bei $A = A'$)

$$(4.3.1)\qquad V(u'Au) = (\mu_4 - 3\sigma^4)\sum a_{tt}^2 + 2\sigma^4\,\mathrm{Sp}\,A^2 .$$

Denn es ist $E\,u'Au = \mathrm{Sp}\,A\,E\,uu' = \sigma^2\,\mathrm{Sp}\,A$ und

$$E(u'Au)^2 = \sum_{t,t';\tau,\tau'} a_{tt'}a_{\tau\tau'}\, Eu_t u_{t'} u_\tau u_{\tau'} =$$

$$= \mu_4 \sum a_{tt}^2 + \sigma^4 \sum_{t\neq t'} (a_{tt}a_{t't'} + a_{tt'}a_{tt'} + a_{tt'}a_{t't})$$

$$= (\mu_4 - 3\sigma^4) \sum a_{tt}^2 + \sigma^4 (2\,\mathrm{Sp}\,A^2 + \mathrm{Sp}^2 A) \quad .$$

4.4. Untergrenze des mittleren quadratischen Fehlers regulärer Schätzer

Der reguläre Schätzfall sei wie folgt definiert:[1)]

(i) Der Vektor y der Beobachtungen $y_1,\ldots,y_T$ entstamme einem Modell, dessen Strukturen durch einen Vektor θ der (funktional unabhängigen) Parameter $\theta_1,\ldots,\theta_m$ vermöge einer Dichtefunktion $f(y;\theta)$ gegeben sind. Der Zulässigkeitsbereich der Parametervektoren sei eine offene Menge $\Theta \subset R^m$. Die Menge $M_\theta := \{y \in R^T \mid f(y;\theta) = 0\}$ hänge von $\theta \in \Theta$ nicht ab.[2)] Für $\theta \in \Theta$ existiere für $j = 1,\ldots,m$ (mit eventueller Ausnahme einer - möglicherweise von θ und j abhängigen y-Menge vom Lebesgue-Maß null) $\partial f(y;\theta)/\partial\theta_j$ und es sei

---

1) Diese Definition erfolgt in Anlehnung an die Voraussetzung des Satzes 1.10 von [Schmetterer], S. 346; vgl. dazu auch [Wilks], S. 348 ff, und [Cramér], S. 479. -
Das nachfolgende Theorem gilt übrigens allgemein, wenn das Modell durch eine Klasse von Verteilungsfunktionen $F(y;\theta)$ gegeben ist, deren zugehörige Maße sämtlich absolut stetig bzgl. eines Maßes μ sind. In den Bedingungen (i) - (iii) tritt μ an die Stelle des Lebesgue-Maßes und die Dichten $f(y;\theta)$ sind als Radon-Nikodym-Dichten bzgl. μ aufzufassen (vgl. dazu [Schmetterer], S. 346).

2) Statt dessen kann man auch verlangen, daß $M(\theta)$ eine - möglicherweise von θ abhängige - Lebesgue-Nullmenge ist (s. [Schmetterer], Satz 1.10 und Fn. 1.23, S. 346).

(4.4.1) $$\frac{\partial}{\partial\theta_j}\int f(y;\theta)\,dy = \int \frac{\partial f(y;\theta)}{\partial\theta_j}\,dy \qquad (j = 1,\ldots,m)\ .$$

(ii) Es sei $E\left[\left(\frac{\partial \log f(y;\theta)}{\partial\theta_j}\right)^2;\theta\right] := \int \left(\frac{\partial \log f(y;\theta)}{\partial\theta_j}\right)^2 f(y;\theta)\,dy < \infty$

$(\theta \in \Theta\ ;\ j = 1,\ldots,m)$.[3]

Ferner sei die Matrix

$$E\left[\frac{\partial \log f(y;\theta)}{\partial\theta}\left(\frac{\partial \log f(y;\theta)}{\partial\theta}\right)';\theta\right]$$

$$:= \left(E\left[\frac{\partial \log f(y;\theta)}{\partial\theta_i}\,\frac{\partial \log f(y;\theta)}{\partial\theta_j};\theta\right]\right)_{i,j=1,\ldots,m}\ ,$$

die hiernach existiert - regulär.

(iii) $\hat{\theta}$ mit den Komponenten $\hat{\theta}_1(y),\ldots,\hat{\theta}_m(y)$ sei Schätzer für θ mit $E(\hat{\theta}_j^2;\theta) := \int \hat{\theta}_j^2\, f(y;\theta)\,dy < \infty$ $(\theta \in \Theta\ ;\ j = 1,\ldots,m)$.

Dann existiert auch $\int \hat{\theta}_i \frac{\partial f(y;\theta)}{\partial\theta_j}\,dy$ [4], und es sei ferner

(4.4.2) $$\frac{\partial}{\partial\theta_j}\int \hat{\theta}_i\, f(y;\theta)\,dy = \int \hat{\theta}_i \frac{\partial f(y;\theta)}{\partial\theta_j}\,dy \quad (\theta \in \Theta\ ;\ i,j = 1,\ldots,m).$$ [5]

---

3) In $M(\theta)$ kann hier und im folgenden Beweis $\partial \log f/\partial\theta_j$ null gesetzt werden.

4) $\int |\hat{\theta}_i \frac{\partial f}{\partial\theta_j}|\,dy = \int |\hat{\theta}_i \sqrt{f}|\,|\frac{\partial \log f}{\partial\theta_j}\sqrt{f}|\,dy \le \left(\int \hat{\theta}_i^2 f\,dy\right)^{1/2}\left(\int \left(\frac{\partial \log f}{\partial\theta_j}\right)^2 f\,dy\right)^{1/2}$

5) Die Voraussetzungen (4.4.1) und (4.4.2) sind für $\theta \in \Theta$ erfüllt, wenn es zu jedem $j = 1,\ldots,m$ ein $h_j(\theta) > 0$ gibt, so daß für $0 < |h| < h_j(\theta)$: $(*)\ |\partial f(y;\theta + he_j)/\partial\theta_j| \le G_j(y)$ ist (ausgenommen eventuell eine von h nicht abhängige y-Menge vom Lebesgue-Maß null), wobei $G_j(y)$ und $\hat{\theta}_i\, G_j(y)$ (Lebesgue-) integrierbar sind. $(j,i = 1,\ldots,m)$ (e_j ist der j-te Einheitsvektor; s. dazu [Cramér], S. 68). - Der Grund, weshalb die y-Ausnahmemenge von $(*)$ nicht von h abhängen darf, liegt darin, daß man die majorisierte Konvergenz von $(f(y;\theta + de_j) - f(y;\theta))/d$ für $d \to 0$

Führt man zur Abkürzung der Schreibweise ein:

$$(\partial f/\partial\theta)' := (\partial f(y;\theta)/\partial\theta_1,\ldots,\partial f(y;\theta)/\partial\theta_m),$$

$$\partial \log f/\partial\theta := (\partial f/\partial\theta)\ /\ f(y;\theta) \quad \text{und}$$

$$\partial E\hat{\theta}/\partial\theta := (\cdot_{ij} = \partial E[\hat{\theta}_i;\theta]/\partial\theta_j)_{i,j=1,\ldots,m},\ V(\frac{\partial \log f}{\partial\theta}) := V[\frac{\partial \log f}{\partial\theta};\theta]\ ,$$

so gilt die folgende Verallgemeinerung eines Theorems von Cramér ([Cramér], S. 480; [Wilks], S. 354) :

Theorem 4.4. Im regulären Schätzfall ist

$$E[(\hat{\theta}-\theta)(\hat{\theta}-\theta)';\theta] \geq \frac{\partial E\hat{\theta}}{\partial\theta}\left[V\left(\frac{\partial \log f}{\partial\theta}\right)\right]^{-1}\left(\frac{\partial E\hat{\theta}}{\partial\theta}\right)' . \tag{4.4.3}$$

Beweis: Nach (iii) existiert $E\hat{\theta} = E[\hat{\theta};\theta] = \int \hat{\theta}\, f(y;\theta)dy$. Ebenso existiert $\partial E\hat{\theta}/\partial\theta$ und es ist $\partial E\hat{\theta}/\partial\theta = \int \hat{\theta}\left(\frac{\partial f}{\partial\theta}\right)' dy$; weiter ist wegen $\int f\,dy = 1$ nach (4.4.1) $\int \frac{\partial f}{\partial\theta}\,dy = 0$, also auch $\int \theta\left(\frac{\partial f}{\partial\theta}\right)' dy = 0$ und somit

$$\frac{\partial E\hat{\theta}}{\partial\theta} = \int(\hat{\theta}-\theta)\left(\frac{\partial f}{\partial\theta}\right)' dy = \int(\hat{\theta}-\theta)\left(\frac{\partial \log f}{\partial\theta}\right)' f\,dy.$$

Mit Vektoren c und v folgt

$$v'\,\frac{\partial E\hat{\theta}}{\partial\theta}\,c = \int v'(\hat{\theta}-\theta)\sqrt{f}\cdot\left(\frac{\partial \log f}{\partial\theta}\right)' c\,\sqrt{f}\,dy$$

Fortsetzung S. 104,5

benötigt; dieser Ausdruck ist gleich $\partial f(y;\theta+\vartheta d\,e_j)$ $(0<\vartheta<1)$ und es genügt, hierfür eine integrierbare Majorante zu haben; jedoch hängt ϑ und damit $h = \vartheta d$ i.d.R. von y ab. - Die angegebene Voraussetzung schließt allerdings aus, daß die y-Menge, für die $\partial f(y;\theta)/\partial\theta_j$ nicht existiert, von θ abhängt.

Anwendung der Schwarzschen Ungleichung liefert

$$(c'(\frac{\partial E\hat{\theta}}{\partial\theta})'v)^2 \le v'E(\hat{\theta}-\theta)(\hat{\theta}-\theta)'v \cdot c'E\ \frac{\partial \log f}{\partial\theta}\left(\frac{\partial \log f}{\partial\theta}\right)' c.$$

Weiter ist $0 = \int \frac{\partial f}{\partial\theta} dy = \int \frac{\partial \log f}{\partial\theta} f\ dy = E\left[\frac{\partial \log f}{\partial\theta}; \theta\right]$, d.h.

$$V(\frac{\partial \log f}{\partial\theta}) = E\ \frac{\partial \log f}{\partial\theta}\left(\frac{\partial \log f}{\partial\theta}\right)'$$

und somit bei festem v $\quad v'E(\hat{\theta}-\theta)(\hat{\theta}-\theta)'v \ge (c'\left(\frac{\partial E\hat{\theta}}{\partial\theta}\right)' v)^2 / c'V(\frac{\partial \log f}{\partial\theta})c$

für jedes $c \neq 0$. Anwendung von (4.1.7) mit $d = \left(\frac{\partial E\hat{\theta}}{\partial\theta}\right)' v$ liefert

nun die Behauptung.

Bezeichnet $b(\hat{\theta}) := E\ \hat{\theta} - \theta$ die Verzerrung des Schätzers, so hat man auch $\frac{\partial E\hat{\theta}}{\partial\theta} = I + \frac{\partial b(\hat{\theta})}{\partial\theta}$. Für unverzerrte Schätzer liefert das Theorem somit in der Tat die bekannte Cramér-Rao-Untergrenze.

Es ist klar, daß man im Prinzip genauer schätzen kann, wenn weniger (unabhängige) Parameter zu schätzen sind. Diesen Sachverhalt drückt aus das folgende

Korollar 4.4. Ist θ_2 Teilvektor von θ, so gilt (bei $\theta' = [\theta_1', \theta_2']$) im regulären Schätzfall

$$(4.4.4)\quad E(\hat{\theta}_2 - \theta_2)(\hat{\theta}_2 - \theta_2)' \ge \frac{\partial E\hat{\theta}_2}{\partial\theta_2}\ V_{22}^{-1}\left(\frac{\partial E\hat{\theta}_2}{\partial\theta_2}\right)' + \left(\frac{\partial E\hat{\theta}_2}{\partial\theta_1} - \frac{\partial E\hat{\theta}_2}{\partial\theta_2}\ V_{22}^{-1}\ V_{12}'\right) V^{11}\left(\frac{\partial E\hat{\theta}_2}{\partial\theta_1} - \frac{\partial E\hat{\theta}_2}{\partial\theta_2}\ V_{22}^{-1}\ V_{12}'\right)'.$$

Darin ist $V_{22} = V(\partial \log f/\partial\theta_2), V_{12} = E(\partial \log f/\partial\theta_1)(\partial \log f/\partial\theta_2)'$ und $V^{11} = (V_{11} - V_{12}\ V_{22}^{-1}\ V_{12}')^{-1}$ mit $V_{11} = (V(\partial \log f/\partial\theta_1)$.

Beweis: Anwendung von (4.1.1) in (4.4.3). -

4.5. Spektral-Darstellung quasi-stationärer Prozesse

Für $t = 1,\dots,T$ sei $x_t' = (x_{1t},\dots,x_{nt})$ mit reellen x_{it} $(i = 1,\dots,n)$. Mit einer reellen, symmetrischen Matrix K_T^{-1} sei

$$(4.5.1)\qquad M^T(\tau) := \begin{cases} K_T^{-1} \sum_{t=1}^{T-\tau} x_{t+\tau}\, x_t'\, K_T^{-1} & (0 \le \tau \le T-1) \\ (M^T(-\tau))' & (-(T-1) \le \tau \le -1) \\ 0 & (|\tau| \ge T) \end{cases}$$

Lemma 4.5. Die Folge $\{M^T(\tau)\}_\tau = \dots,-1,0,1,\dots$ ist positiv-semidefinit. Ihre Spektral-Verteilungsfunktion $\mathcal{F}_T(\lambda)$ ist überall stetig (und erfüllt $\mathcal{F}_T(\lambda) = 0$ für $\lambda \le -\pi$, $\mathcal{F}_T(\lambda) = M^T(0)$ für $\lambda \ge \pi$).

Beweis: Es sei

$$(4.5.2)\qquad f_T(\lambda) := \begin{cases} \frac{1}{2\pi} \sum_{\tau=-(T-1)}^{T-1} e^{-i\lambda\tau}\, M^T(\tau) & (-\pi \le \lambda \le \pi) \\ 0 & (\pi < |\lambda| < \infty)\ . \end{cases}$$

Mit komplexem n - Vektor c ist

$$\left| \sum_{t=1}^{T} c^*\, K_T^{-1}\, x_t\, e^{-i\lambda t} \right|^2 = \sum_{t,s=1}^{T} c^*\, K_T^{-1}\, x_t\, x_s'\, K_T^{-1}\, c\, e^{-i\lambda(t-s)} =$$

$$= c^*\, K_T^{-1} \{ \sum_{t=1}^{T} x_t x_t' + \sum_{\tau=1}^{T-1} (e^{-i\lambda\tau} \sum_{t=1}^{T-\tau} x_{t+\tau} x_t' +$$

$$+ e^{i\lambda\tau} \sum_{t=1}^{T-\tau} x_t\, x_{t+\tau}') \} K_T^{-1}\, c =$$

$$= c^* \sum_{\tau=-(T-1)}^{T-1} e^{-i\lambda\tau}\, M^T(\tau)\, c \quad ;$$

ferner ist $\left(\sum_{\tau=-(T-1)}^{T-1} e^{-i\lambda\tau}\, M^T(\tau)\right)^* = \{M^T(0) + \sum_{\tau=1}^{T-1}\left[e^{-i\lambda\tau}\, M^T(\tau) + \right.$

$$\left. + e^{i\lambda\tau}(M^T(\tau))'\right]\}^* = \sum_{\tau=-(T-1)}^{T-1} e^{-i\lambda\tau}\, M^T(\tau).$$

Somit ist für jedes λ $f_T(\lambda)$ positiv-semidefinit hermitesch sowie (außer in $\lambda = \pm\pi$) auch stetig.

(4.5.3) $\mathcal{F}_T(\lambda) := \int_{-\infty}^{\lambda} f_T(\nu)\, d\nu$

ist überall stetig und für $\lambda \leq -\pi$ ist $\mathcal{F}_T(\lambda) = 0$. Ferner ist $(\mathcal{F}_T(\lambda))^* = \mathcal{F}_T(\lambda)$ und für $\lambda_1 < \lambda_2$

$c^*(\mathcal{F}_T(\lambda_2) - \mathcal{F}_T(\lambda_1))c = \int_{\lambda_1}^{\lambda_2} c^* f_T(\nu) c\, d\nu \geq 0$; also sind $\mathcal{F}_T(\lambda)$ und die Zuwächse von $\mathcal{F}_T(\lambda)$ positiv-semidefinit hermitesch. Schließlich ist (für die stetige beschränkte Funktion $e^{i\lambda\tau}$)

$$\int_{-\infty}^{\infty} e^{i\lambda\tau}\, d\mathcal{F}_T(\lambda) = \int_{-\pi}^{\pi} e^{i\lambda\tau}\, d\mathcal{F}_T(\lambda) =$$

$$= \int_{-\pi}^{\pi} e^{i\lambda\tau} f_T(\lambda) d\lambda = \frac{1}{2\pi} \sum_{r=-(T-1)}^{T-1} M^T(r) \int_{-\pi}^{\pi} e^{i\lambda(\tau-r)} d\lambda = M^T(\tau)\ ,$$ denn

es ist $\int_{-\pi}^{\pi} e^{i\lambda(\tau-r)}\, d\lambda = \delta_{r\tau} \cdot 2\pi$. Somit ist auch für $\lambda \geq \pi$

$\mathcal{F}_T(\lambda) = \mathcal{F}_T(\pi) = M^T(0)$.

Die Positiv-Semidefinitheit der Folge $\{M^T(\tau)\}$ ergibt sich nun wie folgt: Sie ist gezeigt, wenn für jedes $K = 1,2,\ldots$ die Matrix $S_K = \left[M^T(r-s)\right]_{r,s=1,\ldots,K}$ (mit $M^T(r-s)$ im Zeilenblock r, Spaltenblock s) positiv-semidefinit ist. Mit entsprechender Unterteilung des Vektors c ($c' = [c_1', \ldots, c_K']$) ist

$$c^* S_K c = \sum_{r,s=1}^{K} c_r^* M^T(r-s) c_s = \sum_{r,s=1}^{K} c_r^* \int_{-\pi}^{\pi} e^{i\lambda(r-s)} d\mathcal{F}_T(\lambda)\, c_s =$$

$$= \int_{-\pi}^{\pi} \sum_{r,s=1}^{K} e^{i\lambda(r-s)} c_r^* f_T(\lambda) c_s\, d\lambda.$$

$f_T(\lambda)$ besitzt eine Zerlegung $f_T(\lambda) = m^*(\lambda)\, m(\lambda)$, woraus

$$c^* S_K c = \int_{-\pi}^{\pi} \Big| \sum_{s=1}^{K} e^{-i\lambda s} m(\lambda)\, c_s \Big|^2 \, d\lambda \geq 0 \quad \text{folgt.}$$

Theorem 4.5.1. Ist x_t für alle $t = 1,2,\ldots$ definiert und gilt

(4.5.4) $$\lim_{T\to\infty} M^T(\tau) = M(\tau) \qquad (\tau = 0,\pm 1,\pm 2,\ldots)\,,$$

so ist die Folge $\{M(\tau)\}_{\tau = \ldots,-1,0,1,\ldots}$ positiv-semidefinit, und für ihre Spektralverteilungsfunktion $\mathcal{F}(\lambda)$ gilt mit $\mathcal{F}_T(\lambda)$ von (4.5.3):

(4.5.5) $$\lim_{T\to\infty} \mathcal{F}_T(\lambda) = \mathcal{F}(\lambda) \qquad \text{an jeder Stetigkeitsstelle von } \mathcal{F}(\lambda)$$

(sowie $\mathcal{F}(\lambda) = 0$ für $\lambda < -\pi$, $\mathcal{F}(\lambda) = M(0)$ für $\lambda \geq \pi$).

Beweis: Für $c \neq 0$ ist $F_{c,T}(\lambda) = c^* \mathcal{F}_T(\lambda) c$ eine nichtnegative, monoton nichtabnehmende, rechtsstetige Funktion und hat eine von T unabhängige obere Schranke; denn $F_{c,T}(\lambda) \leq c^* \mathcal{F}_T(\pi) c = c^* M^T(0) c$ und dies ist, da $M^T(0)$ konvergiert, beschränkt. Die Folge $\{F_{c,T}(\lambda)\}_{T=1,2,\ldots}$ hat eine Teilfolge, die gegen eine nichtnegative, beschränkte, monoton nichtabnehmende und rechtsstetige Funktion $F_c(\lambda)$ an jeder Stetigkeitsstelle derselben konvergiert. Zum Beweis schreibt man die rationalen Zahlen als Folge $r_1, r_2, \ldots$ hin und bestimmt sukzessive eine Teilfolge, die für $\lambda = r_1$ bzw. $\lambda = r_2$ usw. konvergiert, wobei die für $\lambda = r_{k+1}$ bestimmte Teilfolge ihrerseits Teilfolge der für $\lambda = r_k$ bestimmten ist; die Diagonalfolge hat dann die gewünschte Eigenschaft ([Cramér], S.81 f.). Läßt man c die endlich vielen Vektoren e_j $(j = 1,\ldots,n)$, $e_j + e_k$ $(j,k = 1,\ldots,n;\ j \neq k)$ und $e_j + ie_k$ $(j,k = 1,\ldots,n;\ j \neq k)$ annehmen, so läßt sich nach demselben Beweismuster zeigen, daß es eine Indexfolge T_s gibt, so daß für jeden dieser Vektoren c mit einer Funktion $F_c(\lambda)$ mit den genannten Eigenschaften $F_{c,T_s}(\lambda) \xrightarrow[s\to\infty]{} F_c(\lambda)$ an jeder Stetigkeitsstelle von $F_c(\lambda)$ gilt.

Man definiert nun $\mathcal{F}(\lambda)$ als Matrix mit den Elementen

$$(4.5.6)\quad \begin{cases} F_{jj}(\lambda) = F_{e_j}(\lambda) & (j=1,\dots,n) \\ F_{jk}(\lambda) = \tfrac{1}{2}\{F_{e_j+e_k}(\lambda) - F_{e_j}(\lambda) - F_{e_k}(\lambda)\} + \\ \qquad + \tfrac{i}{2}\{F_{e_k+ie_j}(\lambda) - F_{e_j}(\lambda) - F_{e_k}(\lambda)\} & (j,k=1,\dots,n;\ j\neq k). \end{cases}$$

Offensichtlich ist $\mathcal{F}(\lambda)$ rechtsstetig; weiterhin ist $\mathcal{F}(\lambda) = 0$ für $\lambda < -\pi$. Da $F_{e_j,T} = F_{jj,T}$, $F_{e_j+e_k,T} = F_{jj,T} + F_{kk,T} + F_{jk,T} + F_{kj,T}$ $(j\neq k)$, $F_{e_k+ie_j,T} = F_{jj,T} + F_{kk,T} + i(F_{kj,T} - F_{jk,T})$ $(j\neq k)$ ist, erfüllen die Elemente $F_{jk,T}(\lambda)$ von $\mathcal{F}_T(\lambda)$ ebenso die Gleichungen (4.5.6), wenn man dort nur jeweils mit T indiziert. An jeder gemeinsamen Stetigkeitsstelle λ aller F_c mit den angeführten Vektoren c gilt somit $\mathcal{F}_{T_s}(\lambda) \longrightarrow \mathcal{F}(\lambda)$. Dies gilt folglich für jede Stetigkeitsstelle λ von $\mathcal{F}$; denn für eine solche folgt aus $h\to 0$ bei $\Delta F_\cdot := F_\cdot(\lambda+h) - F_\cdot(\lambda)$: $\Delta F_{e_j} = \Delta F_{jj} \longrightarrow 0$ $(j=1,\dots,n)$ und also für $j\neq k$ $(j,k=1,\dots,n)$ $\tfrac{1}{2}\Delta F_{e_j+e_k} + \tfrac{i}{2}\Delta F_{e_k+ie_k} \longrightarrow 0$, d.h. $\Delta F_{e_j+e_k} \longrightarrow 0$, $\Delta F_{e_k+ie_j} \longrightarrow 0$. Daß $\mathcal{F}$ und die Zuwächse von $\mathcal{F}$ positiv-semidefinit hermitesch sind, ergibt sich nun durch Grenzübergang aus den entsprechenden Eigenschaften von $\mathcal{F}_{T_s}$ und aus der Rechtsstetigkeit von $\mathcal{F}$. Weiterhin gilt für die beschränkte stetige Funktion $e^{i\lambda\tau}$

$$\int_{\pi}^{\pi} e^{i\lambda\tau}\, d\mathcal{F}(\lambda) = \lim_{s\to\infty} \int_{-\pi}^{\pi} e^{i\lambda\tau}\, d\mathcal{F}_{T_s}(\lambda) = \lim_{s\to\infty} M^{T_s}(\tau) = M(\tau) \quad (\tau=0,\pm1,\dots)^{1)}$$

---

1) Da für $c=e_j$ $(j=1,\dots,n)$, $c=e_j+e_k$ bzw. $c=e_k+ie_j$ $(j\neq k;\ j,k=1,\dots,n)$ $F_{c,T_s}(\lambda) \longrightarrow F_c(\lambda)$ an jeder Stetigkeitsstelle von F_c, gilt für diese c: $\int_{-\pi}^{\pi} e^{i\lambda\tau} dF_{c,T_s}(\lambda) \longrightarrow \int_{-\pi}^{\pi} e^{i\lambda\tau} dF_c(\lambda)$; sodann benutze man (4.5.6). Das Integral ist über das abgeschlossene Intervall zu nehmen.

(und also auch $M(0) = \int_{-\pi}^{\pi} d\mathcal{F}(\lambda) = \mathcal{F}(\pi) - \mathcal{F}(-\pi) + (\mathcal{F}(-\pi) - \mathcal{F}(-\pi-0)) = \mathcal{F}(\pi)$ und somit $\mathcal{F}(\lambda) = M(0)$ für $\lambda \geq \pi$).

$\mathcal{F}(\lambda)$ ist also Spektralverteilungsfunktion der (nach (4.5.4) und Lemma 4.5. positiv-semidefiniten) Folge $\{M(\tau)\}$.

Schließlich ist noch (4.5.5) zu zeigen. Es gelte für eine Stetigkeitsstelle λ_o von $\mathcal{F}$ $\mathcal{F}_T(\lambda_o) \not\longrightarrow \mathcal{F}(\lambda_o)$, etwa $F_{j_o k_o, T}(\lambda_o) \not\longrightarrow F_{j_o k_o}(\lambda_o)$. Die beschränkte Folge $\{F_{j_o k_o, T}(\lambda_o)\}$ hat dann eine Teilfolge $F_{j_o k_o, T_r}(\lambda_o) \longrightarrow a \neq F_{j_o k_o}(\lambda_o)$. Nach dem Bewiesenen gibt es eine Teilfolge T_k der Indexfolge T_r und eine Matrixfunktion $\mathcal{G}(\lambda)$ mit:

a) $\mathcal{G}(\lambda)$ ist rechtsstetig

b) $\mathcal{G}(\lambda)$ und die Zuwächse von $\mathcal{G}(\lambda)$ sind positiv-semidefinit hermitesch

c) $\mathcal{G}(\lambda) = 0$ für $\lambda < -\pi$; $\mathcal{G}(\lambda) = M(0)$ für $\lambda \geq \pi$

d) $\int_{-\infty}^{\infty} e^{i\lambda\tau} d\mathcal{G}(\lambda) = \int_{-\pi}^{\pi} e^{i\lambda\tau} d\mathcal{G}(\lambda) = M(\tau) \qquad (\tau = 0, \pm 1, \ldots)$

e) $\mathcal{F}_{T_k}(\lambda) \longrightarrow \mathcal{G}(\lambda)$ an jeder Stetigkeitsstelle von $\mathcal{G}$.

Die Eigenschaften a) - d) bestimmen aber eindeutig die Spektralverteilungsfunktion der Folge $\{M(\tau)\}$ (vgl. [Anderson], S.390 u. S.383). Daher muß $\mathcal{G}(\lambda) = \mathcal{F}(\lambda)$ gelten $(-\infty < \lambda < \infty)$, d.h. es müßte nach e) insbesondere $F_{j_o k_o}(\lambda_o) = \lim_{k \to \infty} F_{j_o k_o, T_k}(\lambda_o) = a$ sein. Das ist ein Widerspruch.

Theorem 4.5.2. $\mathcal{F}(\lambda)$ sei eine rechtsstetige (n x n)-Matrixfunktion mit positiv-semidefiniten hermiteschen Zuwächsen $(-\infty < \lambda < \infty)$. Dann hat $\mathcal{F}(\lambda)$ bzgl. des durch $\sum_{j=1}^{n} F_{jj}(\lambda)$ definierten (vollständigen) Intervallmaßes μ eine Dichte $f(\lambda)$ (d.h. $f(\lambda)$ ist positiv-

semidefinit hermitesch μ - f.ü. und es gilt $\int_A d\mathcal{F}(\lambda) = \int_A f(\lambda)\, d\mu$ für jede μ-meßbare Menge A).

Beweis: Für jeden n-Vektor c ist $F_c(\lambda) := c^* \mathcal{F}(\lambda) c$ rechtsstetig mit reellen nichtnegativen Zuwächsen, bestimmt also eindeutig ein (endliches bzw. σ-finites)[1] vollständiges (Intervall-) Maß μ_c über $\mathbb{R}$; die μ_c-meßbaren Mengen sind genau die Mengen A = B + N, worin B Borelmenge und N Teilmenge einer Borelschen μ_c-Nullmenge ist. (s. [Richter], S.42 und I, §4). Insbesondere gilt dies für die Diagonalelemente $F_{jj}(\lambda) = F_{e_j}(\lambda)$ von $\mathcal{F}(\lambda)$; die diesbezüglichen vollständigen Maße seien mit μ_j bezeichnet, die Systeme der μ_j-meßbaren Mengen mit $\mathfrak{M}_j$ $(j = 1,\ldots,n)$.

μ sei das durch die Funktion $\sum_{j=1}^{n} F_{jj}(\lambda)$ bestimmte vollständige Maß, $\mathfrak{M}$ das System der μ-meßbaren Mengen. Es gilt dann $\mathfrak{M} \subset \bigcap_{j=1}^{n} \mathfrak{M}_j$ und $\mu(A) = \sum_{j=1}^{n} \mu_j(A)$ für $A \in \mathfrak{M}$.[2]

Jedes Maß μ_c ist absolut stetig bzgl. μ: Da für $h > 0$ $\Delta\mathcal{F} = (\Delta F_{jk}) := \mathcal{F}(\lambda + h) - \mathcal{F}(\lambda)$ positiv-semidefinit hermitesch ist, gilt für $j \neq k$ $|\Delta F_{jk}|^2 \leq \Delta F_{jj}\, \Delta F_{kk}$. Somit gilt für $\Delta F_c := F_c(\lambda + h) - F_c(\lambda)$

$$\Delta F_c = c^* \Delta \mathcal{F} c = \sum_{j,k} \bar{c}_j c_k \Delta F_{jk} \leq \sum_{j,k} |c_j||c_k||\Delta F_{jk}| \leq \sum_{j,k} |c_j||c_k| \sqrt{\Delta F_{jj}\, \Delta F_{kk}} =$$

$$= \Big(\sum_j |c_j| \sqrt{\Delta F_{jj}}\Big)^2 \leq |c|^2 \sum_j \Delta F_{jj}.$$

---

1) Für beschränktes $\mathcal{F}$ ist μ_c endlich, andernfalls σ – finit.

2) Für (links halboffene endliche) Intervalle I gilt $\mu(I) = \sum \mu_j(I)$. Wegen der Eindeutigkeit der Erweiterung des Maßes auf das System aller Borelmengen stimmen auf diesem μ und $\sum \mu_j$ überein. Weiter ist $A \in \mathfrak{M}$ von der Gestalt A = B+N, worin B Borelmenge und N Teilmenge einer Borelschen μ-Nullmenge B' ist; aus $0 = \mu(B') = \Sigma \mu_j(B')$ folgt aber $\mu_j(B') = 0$, also $N \in \mathfrak{M}_j$, $A \in \mathfrak{M}_j$, $\mu_j(N) = 0$ $(j=1,\ldots,n)$ und damit $\mu(A) = \sum \mu_j(A)$.

Für links halboffene Intervalle I ist damit $\mu_c(I) \leq |c|^2\mu(I)$, für $\vartheta \in \mathcal{J}$, dem Mengenkörper der endlichen Summen solcher Intervalle, folglich $\mu_c(\vartheta) \leq |c|^2\mu(\vartheta)$ und daher für Summen $S = \sum \vartheta_r$ mit abzählbar vielen Summanden $\vartheta_r \in \mathcal{J}$ ebenso $\mu_c(S) \leq |c|^2\mu(S)$. Ist nun $N \in \mathcal{N}$ mit $\mu(N) = 0$, so gibt es ([Richter], (4.17), S.32) zu jedem $\varepsilon > 0$ ein $S = \sum \vartheta_r$, $\vartheta_r \in \mathcal{J}$ $(r = 1,2,\ldots)$, mit $S \supset N$, $\mu(S) < \varepsilon$ und also $\mu_c(S) < |c|^2\varepsilon$. Damit hat N das bzgl. F_c definierte äußere Maß null und ist μ_c-meßbar mit $\mu_c(N) = 0$ ([Richter], (4.6), S.28; (4.10), S.29, S. 32 o.). - Weiterhin ist damit jede μ-meßbare Menge für jedes c μ_c-meßbar.

Für $c = e_j$ $(j = 1,\ldots,n)$ bzw. $c = e_j + e_k$ $(j,k = 1,\ldots,n;\ j \neq k)$ bzw. $c = e_j + ie_k$ $(j,k = 1,\ldots,n;\ j \neq k)$ sei $f_c(\lambda)$ die Radon-Nikodym-Dichte von $F_c(\lambda)$ bzgl. μ. Man setze

$$(4.5.7)\quad \begin{cases} f_{jj}(\lambda) = f_{e_j}(\lambda) & (j = 1,\ldots,n) \\ f_{jk}(\lambda) = \tfrac{1}{2}\{f_{e_j+e_k}(\lambda) - f_{jj}(\lambda) - f_{kk}(\lambda)\} + \\ \qquad + \tfrac{i}{2}\{f_{e_k+ie_j}(\lambda) - f_{jj}(\lambda) - f_{kk}(\lambda)\} & (j,k = 1,\ldots,n;\ j \neq k). \end{cases}$$

Mit $f(\lambda) := (f_{jk}(\lambda))_{j,k=1,\ldots,n}$ ist dann unter Verwendung von (4.5.6) für jedes $A \in \mathcal{N}$ $\int_A dF(\lambda) = \int_A f(\lambda)\, d\mu$. Für jedes $A \in \mathcal{N}$ ist weiterhin

$\int_A f^*(\lambda)\, d\mu = \int_A dF^*(\lambda) = \int_A dF(\lambda) = \int_A f(\lambda)\, d\mu$ und somit (s. [Richter], (2.20), S.176) $f^*(\lambda) = f(\lambda)$ μ - f.ü. Schließlich ist für $A \in \mathcal{N}$

$\int_A c^* f(\lambda) c\, d\mu = c^* \int_A dF(\lambda)\, c = \int_A dF_c(\lambda) \geq 0$, also (s. [Richter], (2.17), S.175) für festes c $c^* f(\lambda) c \geq 0$ μ - f.ü. Für jeden der abzählbar vielen Vektoren c_r mit lauter rationalen Koordinaten sei N_r die Ausnahme-Nullmenge, in der $c^* f(\lambda)\, c \geq 0$ nicht gilt $(r = 1,2,\ldots)$. Dann ist

$\mu(\bigcup_{r=1}^{\infty} N_r) = 0$ und in $\mathbb{R} - \bigcup_{r=1}^{\infty} N_r$ $\quad c_r^* f(\lambda)\, c_r \geq 0 \quad (r = 1,2,\ldots)$;

somit ist dort überhaupt für jedes c $\quad c^* f(\lambda) c \geq 0$, d.h. $f(\lambda)$ positiv-semidefinit. -

Literaturverzeichnis

Anderson, T.W.: The Statistical Analysis of Time Series. New York 1971

Cramér, H.: Mathematical Methods of Statistics. Princeton 1946

Doob, J.L.: Stochastic Processes. New York 1953

Grenander, U. und Rosenblatt, M.: Statistical Analysis of Stationary Time Series. New York 1957

Hannan, E.J. und Nicholls, D.F.: The Estimation of Mixed Regression, Autoregression, Moving Average and Distributed Lag Models. Econometrica 4o(1972),529-547

Kadiyala, K.R.: A Transformation Used to Circumvent the Problem of Autocorrelation. Econometrica 36(1968), 93-96

Loève, M.: Probability Theory. Princeton 1955

Malinvaud, E.: I Statistical Methods of Econometrics, Amsterdam 1966
II " " " " 1970

Rao, C.R./Mitra S.K.: Generalized Inverse of Matrices and Its Applications. New York 1972

Richter, H.: Wahrscheinlichkeitstheorie. Berlin 1966

Schmetterer, L.: Einführung in die mathematische Statistik. Wien 1966

Schönfeld, P.: Methoden der Ökonometrie, Bd. I. Berlin 1969

Schulmann, H.: Einige Zeitreihenprobleme der Makroökonometrie. Diss., Saarbrücken 1964

Silvey, S.D.: Multicollinearity and Imprecise Estimation. Journal of the Royal Statistical Society, Series B, 31(1969), 539-552

Watson, G.S.: Linear Least Squares Regression. Ann. Math. Stat. 38, 1679-1699

Wilks, S.S.: Mathematical Statistics. New York 1962